44th Annual Conference of the International Military Testing Association

Ottawa, Canada
22-24 October 2002
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IMTA BY-LAWS</strong></td>
<td>vii</td>
</tr>
<tr>
<td><strong>IMTA 2002 STEERING COMMITTEE</strong></td>
<td>xi</td>
</tr>
<tr>
<td><strong>2002 IMTA OTTAWA ORGANIZING COMMITTEE</strong></td>
<td>xiii</td>
</tr>
<tr>
<td><strong>STEERING COMMITTEE MINUTES</strong></td>
<td>xiv</td>
</tr>
<tr>
<td><strong>PAPERS</strong></td>
<td></td>
</tr>
<tr>
<td>Opening Address by Lieutenant-General George Macdonald, Vice Chief of the Defence Staff</td>
<td>1</td>
</tr>
<tr>
<td>Address by Lieutenant General Christian Couture, Assistant Deputy Minister (Human Resources – Military)</td>
<td>4</td>
</tr>
<tr>
<td>Keynote Address by Yvan Clermont: Essential Skills and Labour Markets: Implications for Human Resources Management in the Military</td>
<td>8</td>
</tr>
<tr>
<td>Curnow, C.K.; Belanich, J.; and Freeman, M.W.  Verifying Learners and Test Takers in Distance Learning (DL) Environments</td>
<td>22</td>
</tr>
<tr>
<td>Waugh, G.W.; Putka, D.J.; and Sager, C. E.  Development And Validation Of a U.S. Army Situational Judgment Test.</td>
<td>28</td>
</tr>
<tr>
<td>Putka, D. J.; Waugh, G.W., and Knapp, D.J.  Pay Grade Differences In The Functioning Of The Situational Judgment Test.</td>
<td>37</td>
</tr>
<tr>
<td>Branson, R.K.  A Model for the Concurrent Design of Performance.</td>
<td>46</td>
</tr>
<tr>
<td>Klammer, J.; and Tanner, L.  The Selection of Military Police: An Assessment Centre Approach.</td>
<td>52</td>
</tr>
<tr>
<td>Rumsey, M.G.  Army Selection Research: Past, Present and Future.</td>
<td>62</td>
</tr>
<tr>
<td>Alderton, D.L.  Revitalization of Navy Enlisted Selection and Classification.</td>
<td>71</td>
</tr>
<tr>
<td>Mottern, J.A.; White, M.A.; and Alderton, D.L. 1st Watch on the First Term of Enlistment.</td>
<td>76</td>
</tr>
</tbody>
</table>
Daley, K.E.; Dewar, M.E.  Engineering And Computing Workforce Trends In Dnd And Cf. 82

Vanderpool, M.A.  The Cross Cultural Adaptability Scale: A Psychometric Evaluation 93

Wheeler, J.  Redeveloping the selection process for the Royal Australian Air Force Officer applicants. 105

Stanley, P.P.  Audit of the United States Air Force Promotion Testing Program 113

Bennett, N.  Mental Health Support Following Critical Incidents: The Australian Defence Force Model for Acute Intervention 120

McConne, A.; and Gibbons, P.K.  New Zealand Defence Force On-Going Attitude Survey. 140

Johnston, B.F.; Brown, K.J.; Cole, A.K., and Aggarwal, S. Measuring Organizational Climate in the Canadian Forces. 153

Heffner, T.S.; and Knapp, D.J.  Nco Leadership Tool: Research Approach and Measures 161

White, L.A.  A Quasi-Ipsative Temperament Measure for Assessing Future Leaders. 169

Kilcullen, R.  Using U.S. Army Special Forces Biodata Measures to Predict 21st Century Non-Commissioned Officer Job Performance 176

Boerstler, R.E.; and Kammrath, J.L.  Purposeful Development Of Future Usaf Leaders Through Occupational Analysis. 184

Brady, E.J.  The Automated Survey Generator (AUTOGEN) 188

Sanders, W.R.; and Lickteig, C.W.  Human Command And Control Functional Requirements for Future Systems. 191

Baisden, A.G.; and Rymsza, B. New Directions in Navy Assessment: Developing a Multimedia Prototype. 201

Annen, H.; and Gutknecht, S.  Validity of the Assessment Center for Future Professional Officers (ACABO). 208

Ervin, K.S.  Reducing Advanced Individual Training (Ait) to Less 215
than 26 Weeks.

Collins, N.; and Brown, P. Order Effects of Task Inventory Design. 221

Brady, E.J. The Multi-Skilled Soldier Concept: Considerations for Army Implementation. 235

Brown, K. Career Transitions Within Members of the Canadian Armed Forces 242

Fleck, G. Understanding and Managing Adjustment Disorders in Young Recruits 250

Hetrick, C.; Cooper, G.; Walsh, K.; Smith, R.; Wickes, A.; and Andre, T. Evaluating Communication In A Synthetic Task Environment 257

Halpin, S.M.; McCann, C.; and Pigeau, R. Cross-National Examination of the Aspects of Command. 263

Waldkoetter, R.O. An Appraisal of a Defense Finance Symposium. 272

Jenkins, D. Quality of Life in the Canadian Forces: An Introduction to Definition and Measurement. 283

Held, J.D.; Fedak, G.E.; Crookenden, M.P.; and Blanco, T.A. Test Evaluation for Augmenting the Armed Services Vocational Aptitude Battery. 291

Van de Ven, C.P.H.W. The drop-out rate during training: in line with expectations? 298

Collin, I.A.; and Bender, P. Assessment of the Potential Volume of Releases at the 20 Years of Service Point. 322

Helmick, J.C.; Lee, K.S.; Helba, C.V.; and Perry, S. Using Satisfaction Gaps to Explore Retention Intentions of Military Members. 331

Donohue, J.J. CFAT / AAT Concordance Study: A Preliminary Investigation of the Utility of the AAT as a Proxy for the CFA. 342

Wait, T. Organizational, Social and Demographic Change in the Canadian Forces: 1976 to 2001. 366

Wait, T. Military HR Strategy 2020, Facing the People Challenges of 386
the Future: Consultation Results

Rosen, L.N.; Knudson, K.H.; and Fancher, F. Dimensions of Military Leadership Related to Soldier Well-Being and Readiness. 399

Watson, S.E.; Alderton, D.L.; and Farmer, W.L. Maximizing the Sailor-Rating Match by Considering Aptitude and Interest 411

Farmer, W.L.; Bearden, M.S.; Borman, W.C.; and Hedge, J.W. Navy Psychometrics Of Measures – NAPOMS 416

Wilcove, G.L. Navy Quality of Life Survey Results from Paper and Web Versions 420


Stemate, L.; and Taylor, I. An Innovative Way of Addressing Age-Old Problems in Human Resources 434

Steinberg, A.G. Focus Group Interviews: An Integral Part of Survey Research. 442

Spector, A.; and Gogol, O. Optimal Placement of Soldiers into Multiple Jobs. 446

Sager, C.E.; Putka, D.J.; and Knapp, D.J. Measuring and Predicting Current and Future NCO Performance. 452


Riedel, S.L.; and Karrasch, A. Training Communication Competence for Multinational Teams 467

Olmsted, M.G.; and Underhill, C. Social Presence In Focus Groups: Comparison Of Traditional And Internet Methods. 475

Newell, C.E.; and Rosenfeld, P. Cluster Sampling Versus Stratified Random Sampling Of U.S. Navy Personnel: A Comparison 482

Modjeski, R.B.; and Waldkoetter, R.O. A Cross-cultural Analysis of Distance Learning in the U.S. and India 488

Meijer, M.; Algra, G.E.; and Weerts, J.M.P. Caring for Military 493

Leitão, M.; Thomaz, J.; and Bana e Costa, C. Multimethodology in Practice: An Application in the Portuguese Army Performance Appraisal Model. 511

Heffner, T.S.; and Knapp, D.J. Situational Judgment Test Research Context 528

Mylle, J. Does Experience change Perceived Team Effectiveness in Peace Support Operations? 536

Hess, K.P.; Entin, E.E.; and Miller, P. Simulation Testbed for the Evaluation of Performance (STEP): Innovative Criteria 541

Symposium: New Directions in Navy Selection and Classification Research 550

Chair: William L. Farmer

Michael, P.G.; Hindelang, R.; Farmer, W.L.; Alderton, D.L.; and Watson, S.E. JOIN: Job and Occupational Interest in the Navy 554

Mottern, J.A.; White, M.A.; and Alderton, D.L. Retention and Attrition Among New Recruits 557

Mottern, J.A.; White, M.A.; and Farmer, W.L. Development of the Navy Commitment Scale 560

Hedge, J.W.; Borman, W.C.; Farmer, W.L.; Bearden, R.M.; Ferstl, K.L.; and Kaufman, J.D. Research Directions for the Future of Navy Selection/Classification 562

Blackstone, T.F.; and Williams, P.M. Teams, Strategic Games and Optimal Outcomes 566

Cotton, A.J. The Role of Individual Well Being in Military Capability 570

Levey, M.J.; Cotton, A.J.; and Creamer, M. Screening for Adjustment Difficulties After Peacekeeper missions 576

Cotton, A.J. The Australian Defence Force Mental Health Strategy 581

Borman, W.C.; Hedge, J.W.; Bourne, M.J., Schwerin, M.J.; Kubisiak, 588

van de Ven, C.P.H.W.; and Matser, R.  If You can’t get No Satisfaction: A study into early drop-out during military training in the Royal Netherlands Army

Schmidtchen, D.  The Legacies of an Industrial Age Military Career System: Excessive Job Rotation and Career Tournaments

Short, K.  Behavioral-Measurement Factor for Enlisted Advancement

Lescreve, F.  Why Smart Classification Does Matter

INDEX OF AUTHORS
IMTA BY-LAWS

Article I - Name

The name of the organization shall be the International Military Testing Association (IMTA).

Article II – Purpose

A. Discuss and exchange of ideas concerning the assessment of military personnel

B. Discuss the mission, organization, operations and research activities of associated organizations engaged in military personnel assessment.

C. Foster improved personnel assessment through exploration and presentation of new techniques and procedures for behavioral measurement, occupational analysis, manpower analysis, simulation modeling, training, selection methodologies, survey and feedback systems.

D. Promote cooperation in the exchange of assessment procedures, techniques and instruments among IMTA members and with other professional groups or organizations.

E. Promote the assessment of military personnel as a scientific adjunct to military personnel management.

Article III – Participation

The following categories shall constitute the membership within the IMTA:

A. Primary Membership shall be open to personnel assigned to organizations of the armed services and defense agencies that have been recognized by the IMTA Steering Committee as Member Organizations and whose primary mission is the assessment of military personnel. Representatives from the Member Organizations shall constitute the Steering Committee.

B. Associate Membership shall be open to personnel assigned to military, governmental or other public entities engaged in activities that parallel those of primary membership. Associate members (including prior members, such as retired military or civilian personnel who remain professionally active) shall be entitled to all privileges of the primary members with the exception of membership on the Steering Committee, which may be waived by a majority vote of the Steering Committee.

C. Non-Member Participants represents all other interested organizations or personnel who wish to participate in the annual conference, present papers or participate in...
symposium/panel sessions. Non-Members will not attend the Steering Committee meeting nor have a vote in the association affairs.

Article IV – Dues

No annual dues shall be levied against the members or participants.

Article V – Steering Committee

A. The governing body of the Association shall be the Steering Committee, which will consist of representatives from the Primary Members and those other members as voted by a majority of the Steering Committee. Commanders of the Primary Member organizations will each appoint their Steering Committee Member.

B. The Steering Committee shall have general supervision over the affairs of the Association and shall have responsibility for all activities of the Association. The Steering Committee shall conduct the business of the Association between the annual conferences of the Association by such means of communications as selected by the Chairman.

C. Meeting of the Steering Committee shall be held in conjunction with the annual conference of the Association and at such times as requested by the Chairman. Representation from a majority of the Primary Members shall constitute a quorum.

D. Each member of the Steering Committee shall have one vote toward resolving Steering Committee deliberations.

Article VI – Officers

A. The officers of the Association shall consist of the Chairman of the Steering Committee and a Secretary.

B. The Commander of the Primary Member coordinating the annual conference of the Association shall select the Chairman of the Steering Committee. The term of the Chairman shall begin at the close of the annual conference of the Association and shall expire at the close of the next annual conference. The duties of the Chairman include organizing and coordinating the annual conference of the Association, administering the activities of the IMTA, and the duties customary to hosting the annual meeting.

C. The Chairman shall appoint the Secretary of the Association. The term of the Secretary shall be the same as that of the Chairman. The duties of the Secretary shall be to keep the records of the Association and the minutes of the Steering Committee, to conduct official correspondence for the Association and to insure notice for the annual conference. The Secretary shall solicit nominations for the Harry H. Greer Award.

Article VII – Meetings

44th Annual Conference of the International Military Testing Association
Ottawa, Canada, 22-24 October 2002
A. The association shall hold a conference annually.

B. The Primary Members shall coordinate the annual conference of the Association, either individually or as a consortium. The order of rotation shall be determined by the Steering Committee. The coordinating Primary Members and the tentative location of the annual conference for the following three years shall be announced at each annual conference.

C. The annual conference of the Association shall be held at a time and place determined by the coordinating Primary Member. Announcement of the time and place for the next annual conference will occur at the annual conference.

D. The coordinating Primary Member shall exercise planning and supervision over the program and activities of the annual conference. Final selection of program content shall be the responsibility of the coordinating Primary Member. Proceedings of the annual conference shall be published by the coordinating Primary Member.

E. Any other organization (other than a Primary Member) may coordinate the annual conference and should submit a formal request to the Chairman of the Steering Committee no less than 18 months prior to the date they wish to host.

Article VIII – Committees

A. Committees may be established by vote of the Steering Committee. The Chairman of each committee shall be appointed by the Chairman of the Steering Committee from among the members of the Steering Committee.

B. Committee members shall be appointed by the Chairman of the Steering Committee in consultation with the Chairman of the committee being formed. Committee chairman and members shall serve in their appointed capacities at the discretion of the Chairman of the Steering Committee. The Chairman of the Steering Committee shall be an ex officio member of all committees.

C. All committees shall clear their general plans of action and new policies through the Steering Committee. No committee or committee chairman shall enter into activities or relationships with persons or organizations outside of the Association that extend beyond the approved general plan or work specified without the specific authorization of the Steering Committee.

Article IX – Amendments

A. Amendments of these By-Laws may be made at the annual conference of the Association.

B. Proposed amendments shall be submitted to the Steering Committee not less than 60 days prior to the annual meeting. Those amendments approved by a majority of the
Steering Committee may then be ratified by a majority of the assembled membership. Those proposed amendments not approved by the Steering Committee may be brought to the assembled membership for review and shall require a two-thirds vote of the assembled membership to override the Steering Committee action.

**Article X – Voting**

All members attending the annual conference shall be voting members

**Article XI – Harry H. Greer Award**

A. The Harry H. Greer Award signifies long standing exceptional work contributing to the vision, purpose and aim of the IMTA.

B. Selection Procedures:

1. Prior to June 1st of each year, the Secretary will solicit nominations for the Greer Award from members of the Steering Committee. Prior Greer Award winners may submit unsolicited nominations. Award nominations shall be submitted in writing to the Secretary by 1 July.

2. The recipient will be selected by a committee drawn from the Primary Members and committee members will have attended at least the previous three Association annual conferences.

3. The Chairman of the Award Committee is responsible for canvassing other committee members to review award nominations and reach a consensus on the selection of a recipient of the award prior to the annual conference.

4. The Award Committee selection shall be reviewed by the Steering Committee.

5. No more than one person is to receive the award each year but the Steering Committee may decide not to select a recipient in any given year.

C. The Award is to be presented during the annual conference. The Award is to be a certificate, the text prepared by the officers of the Association and appropriate memorabilia per discretion of the Chairman.

**Article XII – Enactment**

These By-Laws shall be in force immediately upon acceptance by a majority of the assembled membership of the Association.
# IMTA 2002 Steering Committee

<table>
<thead>
<tr>
<th>Country</th>
<th>IMTA Member Organization</th>
<th>Representative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>Australian Defense Force Psychology Organisation</td>
<td>Colonel Tony Cotton</td>
</tr>
<tr>
<td></td>
<td>Campbell Park Offices Department of Defence</td>
<td><a href="mailto:Anthony.Cotton@cdefence.gov.au">Anthony.Cotton@cdefence.gov.au</a></td>
</tr>
<tr>
<td></td>
<td>Canberra, ACT 2600</td>
<td></td>
</tr>
<tr>
<td>*Austria</td>
<td>Ministry of Defence Heerespsychologischerr Dienst</td>
<td>Brig Gen Dr Ernst Frise</td>
</tr>
<tr>
<td></td>
<td>Maria Theresien-Kaaserne A-1130 Vienna</td>
<td><a href="mailto:hpd@bmlv.gv.at">hpd@bmlv.gv.at</a></td>
</tr>
<tr>
<td>Belgium</td>
<td>Section for Accession Policy Directorate General for Human Resources (DGHR-HRP)</td>
<td>LtCol Francois J. Lescreve</td>
</tr>
<tr>
<td></td>
<td>Bruynstraat B-1120 BRUSSELS (N-O-H)</td>
<td><a href="mailto:Lescreve@skynet.be">Lescreve@skynet.be</a></td>
</tr>
<tr>
<td>Belgium</td>
<td>Royal Military Academy of Belgium</td>
<td>Prof Dr. Jacques J. Mylle</td>
</tr>
<tr>
<td></td>
<td>Renaissancelaan, 30 B-1000 BRUSSELS</td>
<td><a href="mailto:Jacques.mylle@psps.rma.ac.ba">Jacques.mylle@psps.rma.ac.ba</a></td>
</tr>
<tr>
<td>Canada</td>
<td>Directorate of Strategic Human Resource Coordination Department of National Defence 101 Col By Drive K1A 0K2 Ottawa, Ontario</td>
<td>Ms. Susan R. Truscott</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="mailto:s.truscott@delbs.ndh.q.dnd.ca">s.truscott@delbs.ndh.q.dnd.ca</a></td>
</tr>
<tr>
<td>Canada</td>
<td>Directorate Human Resource Research and Evaluation Department of National Defence 101 Col By Drive K1A 0K2 Ottawa, Ontario</td>
<td>Col Cheryl Lamerson</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="mailto:lamerson.cd@forces.ca">lamerson.cd@forces.ca</a></td>
</tr>
<tr>
<td>Canada</td>
<td>Directorate of Military Human Resource Requirements Department of National Defence 101 Col By Drive K1A 0K2 Ottawa, Ontario</td>
<td>Mr. Brian Thompson</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="mailto:thompson.br@forces.ca">thompson.br@forces.ca</a></td>
</tr>
<tr>
<td>France</td>
<td>Centre de Recherche du Service de Santé des Armées – Unité de Psychologie 24 Av des Maquis du Grésivaudan BP 87 Cedex F-38702 La Tronche Cedex</td>
<td>Dr. Corinne Cian</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="mailto:CorinneCian@compuserve.com">CorinneCian@compuserve.com</a></td>
</tr>
<tr>
<td>Germany</td>
<td>Federal Ministry of Defence Postfach 1328 D-53003 Bonn</td>
<td>Ms. Wiltraud Pilz</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="mailto:wiltraudpilz@bmvg.bund400.de">wiltraudpilz@bmvg.bund400.de</a></td>
</tr>
<tr>
<td>Singapore</td>
<td>Applied Behavioral Sciences Department Ministry of Defence ABSD, Tower B Defence Technology Towers #16-01, 5 109681 Singapore</td>
<td>Mr. Kian-Chye Ong</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="mailto:okianchy@starnet.gov.sg">okianchy@starnet.gov.sg</a></td>
</tr>
<tr>
<td>Country</td>
<td>Organization</td>
<td>Location</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Sweden</td>
<td>Swedish National Defence College</td>
<td>Karlstad, Sweden</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| The Netherlands | DP 80 GW  
Behavior Science Division Dutch Army                             |                                 | Col Frans Matser
Fransmatser@hotmail.com |
| United Kingdom | Directorate Army Personnel Strategy (Science)  
HQAG (Personnel and Training Command)  
Trenchard Lines, UPAVON  
Wiltshire  
England  
SN9 6BE |                                 | Dr. Sylvia James-Yates
sylvia.james-yates@daps.mod.uk | * |
| United Kingdom | Defense Personnel Security Research Center   
99 Pacific St Suite 455-E  
Monterey, CA 93940 |                                 | Dr. James A. Riedel
riedelja@osd.pentagon.mil | * |
| United States | Navy Education and Training Professional Development and Technology Center | 6490 Saufley Field Road  
32509-5237 Pensacola, FL | Dr. Mike Lentz
Dr-mike.lentz@cnet.navy.mil
Capt Gary Dye
capt-gary.dye@cnet.navy.mil | * |
| United States | US Air Force Occupational Measurement Squadron | 1550 5th Street East  
Randolph AFB, TX 78150-4449 | Mr. Monty Stanley
Paul.Stanley@randolph.af.mil | * |
| United States | US Air Force Personnel Center                                                |                                 | Mr Kenneth Schwartz
Kenneth.Schwartz@afpc.randolph.af.mil |
| United States | US Coast Guard, Occupational Standards                                       | 2100 Second St SW  
Washington, DC 20593-0001 | Ms. Mary L. Norwood
mnorwood@comdt.uscg.mil | * |
2002 IMTA OTTAWA ORGANIZING COMMITTEE

Colonel Cheryl Lamerson  Chair
Ms Susan Truscott  President
Mr. Brian Thompson  Vice President
Major Cheryl Batchelor  Treasurer
Ms Marie McAdam  Conference Coordinator
Ms Catherine Campbell  Secretary
Captain Glen Smith  Assistant Secretary
Ms Carrie Munro
Major Rob Morrow
Captain Nancy Perron
Corporal Doris Tessier
STEERING COMMITTEE MINUTES

International Military Testing Association (IMTA)

Steering Committee Meeting

21 October 2002

The Steering Committee met for the 44th IMTA at 1530 hours, 21 October 2002 at the Chateau Laurier Hotel, Ottawa, Canada. Colonel Cheryl Lamerson, Canadian Forces, chaired the meeting. Steering Committee members are listed on the attachment. Those indicated by an * were in attendance at the Steering Committee meeting.

1. REVIEW OF PREVIOUS MINUTES

Minutes from the 2001 could not be located. We need to place these and future SC minutes on the web site to establish a record of our deliberations.

2. CONFERENCE ADMINISTRATION

Ms. Susan Truscott, Canadian Department of National Defence, described what was planned for the conference, its agenda, and social functions. She also gave some information on activities available in the Ottawa area.

3. FINANCIAL REPORT

Major Cheryl Batchelor, Canadian Forces, reported that we had $12,000 Cdn to forward to the USN for the 2003 event and $11,000 Cdn extra at this date.

4. REVIEW OF BYLAWS

Dr. Jacques Mylle, Belgium, suggested that item III A of the bylaws listed all attendees as being involved in “assessment”, but his was not accurate. A recommendation for change will be considered and submitted in the future as a motion.

5. WEB-SITE UPDATE

Mr. Monty Stanley’s (USAF) report on the web-site is attached. He will be soliciting input from IMTA members and web-site visitors in the next months to continue to update the web-site. Mr. Stanley pointed out that keeping annual proceedings on the web-site is not a problem because their size is manageable. However, if PowerPoint presentations are posted to the web-site they absorb a lot
of capacity. He recommended that PowerPoint files only be kept for one year and not accumulated. CARRIED.

6. AFTER ACTION DOCUMENTATION

Mr. Brian Thompson, Department of National Defence, Canada, reported that all documentation used by the Canadians in preparing this year’s conference would be included in a file for the 2003 host.

7. FUTURE CONFERENCES

Dr. Mike Lentz, USN, reported that plans were underway for the 2003 conference to be held at the Hampton Inn, Pensacola, Florida, USA, the first week of November 2003. The conference will be hosted by the US Navy Education and Training, Professional Development and Technology Centre.

Subsequent conferences were also discussed. The 2004 conference is scheduled for Europe, but no location has yet been determined. Two thousand and five (2005) is confirmed for Singapore. Australia volunteered to host in 2007 and it was suggested that 2006 should be somewhere in North America.

8. HARRY H. GREER AWARD

The guidelines for this award were reviewed and nominations discussed. Mr. Monty Stanley was selected as this year’s recipient. Dr. Lentz discussed the presentation for during the IMTA banquet.

9. PROPOSAL TO CHANGE ASSOCIATION NAME

Dr. Jacques Mylle, Belgium, brought forward the issue of whether the title International Military Testing Association is an accurate description of our association or whether it deters and confuses some potential participants. We are willing to consider the possibility of changing the name, but need to find a replacement that reflects our diverse topic areas, research areas and academic specialities. We will solicit ideas on the web-site.

10. STANDING COMMITTEE

Ms. Susan Truscott, Canadian Department of National Defence, brought forward the idea of a permanent position on the Steering Committee to ensure continuity from year to year. It was suggested that a volunteer stay in one position (e.g., Chair or Secretary) for several years, rather than changing each year. It was discussed that Monty Stanley, as Web-Master, has offered considerable continuity. It was decided that a permanent position was not required, but that ensuring that Steering Committee minutes were put on the web-site would be important.
11. OTHER BUSINESS

a. **Indonesian Contact.** Colonel Cheryl Lamerson, Canadian Forces, reported that the Psychology Service of the Indonesian National Army had contacted IMTA about possible membership. They have requested that we send them an official letter telling them about the Association and conference. It was agreed that the 2002 organizing committee would do this.

b. **Legacy Funds.** It was discussed whether a specific amount of legacy funds should be passed on each year for advanced conference planning. It was agreed that the aim of our conferences is not to lose money, and to try to pass some seed money to the following conference. Legacy funds in the range of $10,000 (US dollars) were considered ideal, with a maximum of $15,000 (US dollars).

c. **Letter Announcing IMTA.** The letter from a host 3 Star to 3 Stars of the other Steering Committee nations has proven very positive. It was decided it should continue.

12. ADJOURNMENT

We adjourned at 1730 hours.

Colonel Cheryl D. Lamerson, C.D., Ph.D.
Canadian Forces
Director Human Resources Research and Evaluation
Opening Address

Lieutenant-General George Macdonald

Vice Chief of the Defence Staff

at the International Military Testing Association

44th Annual Conference

Ottawa, Ontario

October 22, 2002

Ladies and Gentlemen:

I’m very pleased to welcome you all on behalf of the Chief of Defence Staff, General Henault. He regrets that he could not join you himself; particularly given the priority he places on personnel issues. He asked me to relay his best wishes for a productive conference. Au nom du Général Henault, Chef d’état-major de la Défense, je vous souhaite la bienvenue au Canada et à cette conférence. Il regrette de ne pas pouvoir être présent ici aujourd’hui. As the resource manager for the Canadian Forces, I too have a particular interest in the work carried out by the International Military Testing Association. For 44 years now, this organization has been providing member countries with cutting edge research on military human resources. Looking at your agenda for the next few days and the calibre of people who will be making presentations, I don’t think it’s premature to say that a lot of important work is going to get done. You will be touching on virtually every personnel issue facing our militaries these days -- from recruitment to retirement and everything in between. And in doing so, you’ll be reflecting the fact that we need, now more than ever, a holistic approach to human resource management in the military. In this regard, I would like to briefly give you a few examples of the initiatives we currently have underway in the Canadian Forces. They reflect our emphasis on one of our five corporate priorities – Putting People First.

Quality of Life

The first area I’d like to talk about is quality of life. A few years ago, a parliamentary committee conducted an in-depth study of quality of life in the Canadian Forces and as a result, made a number of recommendations on how it could be improved. To implement these recommendations, we opened a Quality of Life Project Management
Office, which was so successful that we have actually stood it down. In its place, however, we now have a permanent Directorate of Quality of Life, which will maintain the focus into the future. Among the areas where we’ve made good progress is health care reform. For example, the Canadian Forces Health Protection Program aims to prevent illness and injury. It focuses on taking a more proactive approach to health and well-being. Under this program we are making very good progress in dealing with hazards that are specific to military operations, including environmental hazards. In fact, our first ever health-hazard assessment team went to Afghanistan, where Preventive Medicine Technicians and one of our newly trained Industrial Hygiene Officers conducted a full environmental assessment of the air, water and soil quality, as well as a full health risk assessment. Something else we’re dealing with -- as I’m sure many of you are as well -- is treating members with operational stress injuries, particularly Post Traumatic Stress Disorder or PTSD. To address this problem, we have established Operational Trauma and Stress Support Centres across the country. And we are complementing these with initiatives to strengthen mental health services. In fact, Statistics Canada is conducting a national mental health survey, which includes a sample of Regular and Reserve Force members. This will allow us to identify issues relevant to the military and to conduct comparisons with the rest of the population. But perhaps the biggest quality of life issue comes as a result of our high operational tempo. Over the last year, the Canadian Forces have been heavily involved in the international campaign against terrorism. At our peak, we had almost 2,400 troops deployed in and around Afghanistan. Added to the more than 1,500 members we have in Bosnia, this level of deployment was putting a great deal of pressure on our forces. Our priority now is to examine how we can meet our international commitments and protect Canadians at home while giving our members the time they need for recuperation, leave and training. We don’t yet have all the answers but it is something we are addressing in earnest. Clearly, a major element of the answer is a stable, predictable deployment rate – something we haven’t had for several years.
Professional Development

But our focus on people isn’t just about dealing with problems, whether it’s before or after they occur. We are also working hard on professional development. In the Canadian Forces, we’re doing this by promoting a learning culture at all rank levels. First, it is one way of making sure that our people have the intellectual tools they need to do their jobs. Knowledge and critical thinking skills can be just as important to getting a job done as the right piece of kit. Second, by helping our people further their education, we make it easier for them to pursue their personal career goals. This in turn increases career satisfaction -- and that is good for both recruitment and retention. So whether it's higher education, specialized training or mentoring, we will be making sure that all members of the Canadian Forces have access to professional development opportunities throughout their careers.

Conclusion

My colleague, General Couture, will elaborate further on some of these initiatives. But I would also encourage you to share your own experiences with one another over the next few days. Make the most of your time together. Ask the probing questions and explore new ideas. As a group, you have the collective capacity to develop new ideas, challenge traditional thinking and perhaps even help us chart a new course when it comes to our personnel management. Because as we all know, people are the lifeblood of our militaries. So once again, welcome to Canada and enjoy your conference.
Address

Lieutenant General Christian Couture
Assistant Deputy Minister (Human Resources – Military)

At the 2002 Conference of the
International Military Testing Association

Ottawa, Ontario

22 October, 2002

Good morning, Ladies and Gentlemen. It is my pleasure to be here today and to take part in the opening of your conference. J’espère que vous allez tous profiter de cette opportunité d’apprendre de vos collègues – et aussi de vous amuser dans notre belle ville.

I would like to start my presentation this morning by reading to you the Mission of the Human Resources – Military Group.

“To develop and implement plans, policies and programs to recruit, develop and retain people to effectively support the CF in operations and meet the Defence Mission.”

As the Assistant Deputy Minister in charge of the HR-Military Group, my mandate is not only to react quickly to operational imperatives, but also to plan – to develop Strategic guidance for HR. Strategic guidance must be agile - and must respond to changes in the environment. It is the statement of “how” we will accomplish the HR mission, providing direction for the CF senior leaders to effectively address the major strategic issues that we face. As such, it must be driven by an analysis of the best research available: research on demographics (internal and external to the organization) and social and economic trends, to help us identify the issues that we must be ready to address; research on leadership, on recruiting and selection, on retention, training and
professional development, competencies and occupational analysis, military culture, quality of life and HR systems, to name but a few. Organizations such as yours play an important part in the sharing of such knowledge - IMTA in particular, since it focuses specifically on military HR issues.

Strategic planning should be more than just the programming of future plans. In the past, our strategic HR planning was based on the assumption that the status quo would continue, or at least, that the trends we had seen in the recent past would continue. While this may be a “reasonable” assumption, it tends to produce an “if it ain’t broke, don’t fix it” mentality towards HR practices. Consequently, it’s an approach that can find us unprepared when unforeseen events unfold.

In an attempt to increase our preparedness, therefore, we have recently adopted a scenario planning approach for HR strategic planning. Four different scenarios were developed, that represent a series of plausible future changes – including social, technological, economic, environmental and political changes. These scenarios, should they unfold, would lead to four very different “futures” 20 years from now.

Planning for these possible alternative futures is what provides us with the ability to think creatively and strategically about the future. It gives us the context we need to develop new HR strategies and objectives that will work under all four scenarios (or any number of possible variations).

I would like to briefly explain how this works. Under each future scenario, we have postulated a series of events and social and economic changes (both national and international), for the short term, mid-term and long-term horizons. The long-term horizon takes us up to the year 2020. Each of these scenarios has implications for HR: the type of recruit we would be seeking, the availability of suitable recruits, the selection instruments we would need, the competencies our members would require, the focus of
training, HR policies (such as more flexible Terms of Service, or changes to compensation and benefits), attrition/retention issues, and so on.

Once we have considered these implications, we are in a position to develop strategic options for each scenario, which will carry us through the short, mid and long-term. Based on these options, we are currently developing our strategic HR objectives for the short, mid and long-term – objectives that should ensure that we do not get surprised by a future which may be considerably different from 2002!

Nos objectifs stratégiques en matière de RH, qui sont directement liés aux objectifs fondamentaux de la Stratégie de défense 2020 du Canada, sont établis dans les domaines suivants : le leadership, la planification stratégique, la culture, les communications, la consultation, le maintien en poste, le recrutement, le perfectionnement professionnel, la transition, la santé, le bien-être et les systèmes de gestion des RH.

In the course of developing our Strategic HR objectives, some of the many research questions we will need to answer are:

(1). What are some of the competencies that will be required of our future leaders, who will have to operate in an environment characterized by unpredictable, radical change that often blurs the boundaries between military and domestic operations and national and international commitment?

(2). What changes in leadership style may be needed for our leaders to operate effectively in a flatter organization, incorporating the networked approach to doing business, while retaining the chain of command for oversight?

(2). Comment pouvons-nous évaluer les pratiques, les programmes et les systèmes de RH et intégrer ces paramètres dans le Cadre de mesure du rendement du Ministère, de
telle sorte que le travail puisse être évalué adéquatement du point de vue tant quantitatif que qualitatif? Comment pouvons-nous démontrer la valeur des RH au sein d’une organisation où le «profit » n’est pas le résultat recherché?

(3). How can a knowledge management approach facilitate HR “systems” thinking and improve effectiveness throughout the military HR System?

(4). What is the optimal ratio of distance learning to on-site training courses and how do we encourage greater participation in distance learning?

(5). In our efforts to increase retention, how do we maintain a balance between improving transition assistance to members on release or transfer to the Reserves - yet in such a way that it does not actually encourage attrition?

(6). How do national cultural differences in values and military ethos affect operational effectiveness in joint military operations?

(7). How do we put in place a reward system that encourages strategic planning among our senior officers, in an organization where tenure on the job is generally only 1 to 4 years?

These are but a few of the many HR questions we will need to find answers to in the coming years. In seeking these answers, we must take into consideration the changing demographics, and the implications for employers of the changing values and expectations of youth; these are topics that will be discussed by your guest speaker, Scott Murray, from Statistics Canada. We will also continue to look to organizations such as IMTA and to researchers such as yourselves, to point the way ahead for us. On that note, I wish you every success in your conference.
Essential Skills and Labour Markets: Implications for Human Resources Management in the Military

Yvan Clermont
International Coordinator
Adult Literacy and Lifeskills Survey
Statistics Canada

Telephone: (613) 951-3326
e-mail address: cleryva@statcan.ca
http://www.ets.org/all/

2002 IMTA Conference
Ottawa, Canada
October 22, 2002

This presentation:

I. The Emerging Demographic Challenges

II. The Emerging Skill Levels Challenges
I. The Emerging Demographic Challenges

Five-year growth rates (%) of selected countries (1996-2001)
Total fertility rate, Canada and the United States, 1940-2000

Number of children per woman

Immigration is an increasingly important component of population growth in Canada
Immigrants increasingly from Asia and Middle East; visible minorities concentrated in urban areas

![Immigrant Population by Place of Birth](image)

![Visible Minority Population as a Percentage of Census Metropolitan Areas, 1996](image)

Median Age, G8 countries and Australia

![Median Age Graph](image)
II. The Emerging Skill Levels Challenges
Why Conduct Literacy Surveys or Measure Human Capital?

• Human capital definition:
  – …the knowledge that individuals acquire during their life and use to produce goods, services or ideas in market or non-market circumstances.

• Human capital is linked to:
  – economic success
  – productivity
  – social cohesion
  – various indicators of well-being (health, social exclusion, welfare, crime…)

• Benefits of investing in human capital:
  – …investments in human capital are made in the development of human resources in the expectation of a return; e.g. Productivity, earnings, etc.

Sources of Policy Interests

• Increasing Relative Importance Of Human Capital
  • Technological Change
  • Globalization
  • Relatively Small Youth Cohorts

• Dissatisfaction with Econometric Treatment of Human Capital Theory
• Employer Dissatisfaction with the Quality of Skill Supply
The Skills of Youth

PISA: 15 years old
about to enter the labour market

PISA: What it measures and what it tells us

Reading Literacy
Using, interpreting
and reflecting on
written material
Mathematical literacy
Recognising, formulating and solving mathematical problems in everyday life. Analysing and reasoning and communicating in mathematical terms

Reading Literacy
Using, interpreting and reflecting on written material

Scientific Literacy
Using scientific knowledge, identifying scientific questions, and drawing evidence-based conclusions to understand and make decisions about the natural world
Reading Skills Across Top & G8 countries

Mathematical Skills Across Top & G8 countries
Science Skills Across Top & G8 countries

Percentage of students at each reading proficiency level

The Skills of Adults

IALS/ALL: 16-65 years old
about those in the labour market

Many adults in many countries have weak literacy skills

Percentage of adult population aged 16-65 at each prose literacy level, 1994-1998

- Level 1
- Level 2
- Level 3
- Level 4/5
When comparing Youth (16-25) to older generation (46-55)
Some high-school leavers less likely to have adequate literacy skills

Earnings, education and literacy
Earnings and literacy proficiency, controlling for education and labour force experience

Countries are ranked by the magnitude of the effect parameter associated with literacy proficiency.
Likelihood of receiving employer support for training

Countries are ranked by the odds of the 4th quartile. The statistical difference to the United States is computed for the 4th quartile.

Note: Statistical difference is significant at p < .05.

Military training increasingly occurs at sites outside the traditional classroom through the use of the Internet. Much of this training occurs in the absence of the instructor, both physically and temporally. Although cheating is not currently identified as a problem in the US Army, reliance upon such asynchronous distance learning systems increases the odds of various forms of training compromise, such as enlisting a proxy for test taking in non-proctored, web-based learning environments.

Training compromise and the prevention of compromise in DL training and testing environments have many implications, from the level of confidence that individuals have in the quality of DL programs to issues of military readiness. Just as is the case in the traditional classroom, it is important that DL training programs have mechanisms to ensure that training and testing procedures are not being compromised. How can it be determined whether the student online is the intended learner, particularly during individual testing?

To begin to answer this question, a workshop titled “Training on the Web: Identifying, Authenticating, and Monitoring Learners” was conducted in November 2001 at Carnegie Mellon University. The workshop was sponsored by the Army Training Support Center, which recognized that the increased use of computer technology for distributed learning systems might lead to future problems with online training and testing. There currently is no definitive evidence that compromise during online testing is a problem in the Army. However, the increased use of distributed learning coupled with reports of increased frequency of cheating among high school students (DeWan, 1994; McCabe, 2001) and college students (Argetsinger, 2001) is reason for concern. During the workshop remedies to compromise were sought from experts in a variety of fields including: biometrics (e.g. fingerprint identification and iris scanning), commercial testing centers, computer security, military law, and training. This paper presents research and recommendations on verifying and authenticating learners as a safeguard to training compromise (or cheating) in Internet-based distance learning environments.

Methodology

The workshop included presentations by experts and concluded with a brainstorming session during which the participants generated potential solutions. Thirty-one individuals from industry, academia, and government attended. The presentations
covered the following topics: a) training and testing design, where a course and its assessment can be designed to diminish the possibility of cheating; b) test security, explaining how the Educational Testing Service prevents and detects cheating; c) Public Key Infrastructure (PKI), a system to authenticate and secure transmission of information across the Internet using asymmetric encryption; d) biometrics, the process of identifying people based on their physical, personal, and/or behavioral characteristics; and e) military legal issues regarding training compromise, where cheating is a failure to follow a rule or regulation and, therefore, can be treated as misconduct. After the workshop, an advisory panel met to discuss final recommendations.

Findings and Recommendations

Based on the presentations and discussions, recommendations were developed which are presented here in no particular order of merit. The recommendations are meant to function as general guidelines. The usefulness of any particular solution depends on the training situation. For example, the use of biometrics for identifying a learner might be useful for a course that covers sensitive material or courses that lead to some form of certification, but it may be an excessive measure for a course that is merely a refresher of basic general knowledge. Also, there may be privacy concerns and policies specific to an organization regarding the encoding and storage of individual biometric data.

Affirmative Obligations

Affirmative obligation statements involve presenting a statement that details appropriate and inappropriate behavior during a test. Learners are required to sign the statement or click on a “submit” or “accept” button on the monitor, as in the generic example below. Affirmative obligations assist in establishing that a transgression occurred.

Figure 1. A generic affirmative obligation statement adapted from the U. S. Army.

Biometrics

The Defense Biometrics Management Office described the four steps in the process of using biometrics for identification: capture, process, enroll, verify. The capture process is where the device obtains the biometric data (e.g., fingerprint, iris
image, handwriting sample). The data are then processed and encoded to an easily storable form. This encoded data can also be encrypted to provide a higher level of security. The enrolling procedure occurs the first time an individual’s biometric data are obtained and stored. The verification procedure occurs when an individual’s biometric data are compared to stored data to determine if a match has occurred. The technology was demonstrated during the workshop.

There are some concerns with biometrics. For example, the technologies are not completely foolproof. There are two types of errors that can be made. The first is a false acceptance, when the biometric data of a confederate is accepted. The second is a false rejection, when the correct person is rejected as not matching his/her biometric profile. To lessen the chance for these errors, biometrics can be used in conjunction with a password, or more than one biometric can be used at a time to identify a person. The most common and promising biometric technologies are fingerprints, iris patterns, facial composition, and handwriting.

**Fingerprint identification.** This biometric identifies between 15 and 20 distinct minutia points (e.g., end of a ridge, joining of two ridges) in a single fingerprint, and then the distance and angle between the key points are measured. The fingerprint is thereby encoded as vector measurements between these points. The actual fingerprint is not stored but rather the polygon that connects the distinct characteristics of a fingerprint; thus, the fingerprint image cannot be duplicated from the encoded set of vector measurements, increasing security.

**Iris patterns.** Scanning a person’s iris (the colored part of the eye surrounding the pupil) is a method that is promising, and may be a preferred method in the future. In this procedure, the pattern of radial marks and their relative position in the iris are obtained. The iris pattern remains identifiable, even though the size of the iris changes as the pupil changes due to lighting.

**Facial composition.** The facial composition method identifies points on the face (i.e., eyes, nose, mouth) and then measures the relative distance between these key points. The relative distance and angles between these facial components can be encoded in a similar manner as fingerprints and iris patterns.

**Handwriting analysis.** With current technology, handwriting analysis involves more than just the visual shape of the signature. The speed and pressure used to create the signature are measured in addition to the shape of the signature. The inclusion of speed and pressure make this biometric much more difficult to forge than just the visual aspects of a signature.

There are some other forms of biometrics, such as retinal images, hand geometry, vein pattern on the back of the hand, and ear shape. The technology for these forms of biometrics either are still emerging and/or may not be used due to the invasive nature of the measure. Cost is also a factor; to secure a computer with a fingerprint system can cost approximately $100 and an iris scanning system approximately $250. With further
development, these prices are likely to fall in the future. Of the various biometric technologies, fingerprint scanning is the most mature method based on cost, reliability, and usability. During the workshop, the Center for Identification Technology Research, sponsored by the National Science Foundation, presented on future biometric technology, such as perspiration patterns, template aging, and body odor.

![Fingerprints from the same finger, but taken at different angles. The lines link corresponding minutia points between the fingerprints.](image)

**Figure 2.** Fingerprint scans from the same finger, but taken at different angles. The lines link corresponding minutia points between the fingerprints.

**Proctoring**

There are two types of proctoring: live and virtual. Live proctoring requires that students go to a specific location for monitoring by another person; virtual proctoring involves using technology to monitor students at remote locations.

Live proctoring is expensive but hard to beat for high stakes testing situations where the consequences of cheating and not knowing material are severe. For example, it is imperative that a helicopter mechanic is aptly qualified to identify and repair faulty parts. There are several companies that provide live proctoring services.

For virtual proctoring, the advisory panel recommends using a layered approach depending on the critical nature of the test. Examples of how this layered approach may work are listed below.

- With high stakes tests, video monitoring and a biometric measure such as iris scanning may be used.
- For medium stakes tests, a single biometric measure may be acceptable.
- For low stakes tests, no proctoring measures may be needed.

Alternatively, or in addition to biometrics, students can be asked biographical questions (e.g., last 4 digits of social security number or mother’s maiden name) during testing to help verify their presence. The final virtual proctoring recommendation is to track keystrokes and the web sites visited by the test taker. This can provide evidence if cheating occurs, and can also serve as a deterrent.
Public Key Infrastructure

The use of PKI limits unauthorized access to tests and assures that materials are not altered en route. As PKI encryption becomes more common, the ease of incorporating the technology into distance learning courses will increase.

Test design

The principal test design recommendation is to use performance-based testing, where the student must demonstrate performance proficiency. Successful completion of a performance-based test indicates that the test taker is able to perform the task, and pre-knowledge of test content becomes almost irrelevant. Other recommended design techniques are to randomize test items and use multiple forms. This reduces the utility of answer keys from previous tests.

The final set of test design recommendations involves setting limits. First, appropriate time limits can be set so test takers do not have time to complete the test plus look up answers in reference materials. Additional limits include restricting the number of times that a student can take a test and disabling computer “print/capture” options to reduce the possibilities of sharing test items with others.

Conclusion

It was the advisory panel’s basic assumption and belief that learners generally will “do the right thing.” The solutions, however, are meant to level the playing field, dissuading potential cheaters while not burdening those who never intend to cheat. The purpose of the study was to identify remedies to compromise in Web-based training and testing environments that can be implemented without hindering learning and prior to any problems arising.

The overall recommendation is a layered approach based on the criticality of the test under consideration. In some situations the use of biometrics, live proctoring, and encryption may be warranted, while in other situations only the use of an affirmative obligation statement may be appropriate. The level of security would depend on multiple variables determined by a course administrator and organizational policy.

References


For more information please contact:
Dr. Christina Curnow
Senior Associate
Caliber Associates
10530 Rosehaven St., Suite 400
Fairfax, VA 22030

curnowc@calib.com
phone: (703) 279-6213
DEVELOPMENT AND VALIDATION OF A
U.S. ARMY SITUATIONAL JUDGMENT TEST

Gordon W. Waugh, Ph.D.
Senior Scientist
Human Resources Research Organization
66 Canal Center Plaza, Alexandria, VA 22314
gwaugh@humrro.org

Dan J. Putka, Ph.D.
Research Scientist
Human Resources Research Organization
66 Canal Center Plaza, Alexandria, VA 22314
dputka@humrro.org

Christopher E. Sager, Ph.D.
Manager Assessment Research and Analysis Program
Human Resources Research Organization
66 Canal Center Plaza, Alexandria, VA 22314
csager@humrro.org

INTRODUCTION

In this project, two multiple-choice situational judgment tests were developed: the SJT and the SJT-X. The final SJT test consists of two 24-item forms (one for promotion to the E5 level, the other for promotion to the E6 level). The SJT-X has only three items, but its scenarios (i.e., situation descriptions) average 700 words in length. The development of these SJTs had five unusual aspects. First, the SJT-X situation descriptions were much longer than usual. Second, several scoring algorithms were compared. Third, an automated iterative process was used to decide which items and response options to keep. Fourth, alternative methods of item selection were compared. Fifth, the SJT, and especially the SJT-X, were developed to assess aptitudes believed to be much more important to performance in the future Army. The SJT was also used to examine the leadership styles of junior NCOs.

The SJT items were developed to measure eight aptitudes relevant to being an effective NCO: (a) Directing, Monitoring, and Supervising Individual Subordinates, (b) Training Others, (c) Team Leadership, (d) Concern for Soldiers’ Quality of Life, (e) Cultural Tolerance, (f) Motivating, Leading, and Supporting Individual Subordinates, (g) Relating to and Supporting Peers, and (h) Problem-Solving/Decision Making Skill.

1 This paper is part of a symposium titled Tailoring a Situational Judgment Test to Different Pay Grades presented at the 2002 International Military Testing Association Conference in Ottawa (G.W. Waugh, Chair). The views, opinions, and/or findings contained in this paper are those of the authors and should not be construed as an official U.S. Department of the Army position, policy, or decision.
A sample SJT test item is shown in Figure 1. Each item presented a 2–4 sentence scenario (i.e., description of a problem situation) followed by four possible actions (see Figure 6.1). Soldiers were instructed to indicate (a) which action was most effective and (b) which action was least effective.

<table>
<thead>
<tr>
<th>One of your fellow soldiers feels like he doesn’t have to pitch in and do the work that you were all told to do. What should you do?</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Explain to the soldier that he is part of a team and needs to pull his weight.</td>
</tr>
<tr>
<td>b. Report him to the NCO in charge.</td>
</tr>
<tr>
<td>c. Find out why the soldier feels he doesn’t need to pitch in.</td>
</tr>
<tr>
<td>d. Keep out of it; this is something for the NCO in charge to notice and correct.</td>
</tr>
</tbody>
</table>

Figure 1. Sample SJT test item.

The SJT-X was designed to assess the construct, Knowledge of the Inter-Relatedness of Units. Soldiers use this construct in complex situations. Therefore, the three situation descriptions for the SJT-X are lengthy. Two of the descriptions are a half page long and the third is almost two pages.

DEVELOPING THE ITEM CONTENT

The SJT and SJT-X development had three components: (a) item writing and editing, (b) data collection and analysis, (c) selection of response options, and (d) selection of items for the final test forms. The data collection and analysis took place at various times during the development. The major data collections were a field test ($N = 495$) and a validation ($N = 1,851$).

**Item Writing**

The development of the draft SJT items consisted of several steps: (a) relevant soldier aptitudes were identified, (b) some items were selected from existing situational judgment tests used in two previous Army studies, (c) critical incidents from three previous Army studies were examined for potential use as scenarios in new SJT items (none of these critical incidents were useful for various reasons: they were too vague, did not represent problem situations, or the correct action was obvious), (d) soldiers wrote descriptions of situations they had observed that were related to the relevant constructs, and (e) other soldiers described what they would do in those situations. Throughout the development process, items were edited for style and content. The SJT-X was developed by a single member of the project staff (who was a retired Army officer) using his knowledge of the target dimension.

The SJT and SJT-X were tested in two phases: a field test\(^2\) and a concurrent validation.\(^3\) The field test analyses are based upon 495 soldiers for the SJT and 24

---

\(^2\) This paper refers to various versions of the SJT. Each soldier in the field test completed one of two SJT forms. Form A had 42 items and Form B had 39 items. The two forms had 14 items in common. After selecting the best set of 40 items from the field test (5 items in each of the 8 scales), subsequent field test analyses were done using this shortened 40-item SJT.
soldiers for the SJT-X. The validation analyses are based upon 1,851 soldiers for the SJT and 470 soldiers for the SJT-X. The goals of the field test were to determine which items and response options to retain and which response format to use. Before the field test, senior NCOs were used as Subject Matter Experts (SMEs) to rate the effectiveness of the response options of the SJT items. Based on these ratings, the interrater reliability for the total score (of the 40-item form) was .91 (.40 for 1 rater). The goals of the validation were to finalize the test forms and evaluate their validity and other psychometric characteristics.

The SJT form used in the validation had 40 items. Each item presented a 2–4 sentence scenario (i.e., description of a problem situation) followed by four possible actions (see Figure 1). Soldiers were instructed to indicate (a) which action was most effective and (b) which action was least effective. Each of the eight KSAs was represented by five items. The development of the SJT is described in detail by Knapp et al. (2002). For the field test, soldiers also rated the effectiveness of each response option on a seven-point scale.

Selection of Response Options

Most SJT draft items had more than the four response options desired for the final test forms. Therefore, some response options had to be dropped after pilot testing. The field-test version of the SJT had 67 items; the number of response options ranged from 4–7. The items were split into two overlapping forms (Form A and Form B). Six scoring algorithms were computed for each item. Item scores were computed for all possible combinations of four response options. Thus, a seven-option item had 35 different combinations of four options. Because six scoring algorithms were used, a seven-option item would be scored 210 ways. In all, item scores were computed for 5,682 different combinations of response options.

Because there were too many sets of response options to manually select the best set of response options for each item, this process was automated. For each item, the set of four response options that maximized the reliability of the item’s relevant scale (from among the eight aptitude scales) was retained. The final SJT-X form could have more than four options per item. Thus, only options with poor psychometric characteristics would be dropped from the SJT-X.

FINALIZING THE TEST FORMS

Finalizing the test forms consisted of deciding how the tests would be scored and picking which items to keep.

---

3 Only one 40-item SJT form was used in the validation. After constructing two 24-item forms (3 items in each of the 8 scales) based on the validation data, subsequent validation analyses were done using these shortened 24-item SJTs. One form was for promotion to the E5 level, the other for promotion to the E6 level.

4 A set of SAS (Statistical Analysis Software) macros was written to select the response options.
Selection of the Scoring Algorithm

There are many ways to score an SJT. A total of 10 scoring algorithms were compared in the validation (3 additional ones were compared in the field study). In the end, it was decided to use algorithm 10: The score for an item is computed by subtracting the keyed effectiveness (i.e., the SMEs’ effectiveness rating) of the option selected by the soldier as least effective from the keyed effectiveness of the option selected as most effective.

Table 1 shows the correlations among the 10 scoring algorithms compared in the validation. It appears that the algorithms are measuring very similar things. There does appear to be a difference, however, between the algorithms that give points for identifying the best response (1, 4, and 8) and those that give points for identifying the worst response (2, 5, and 9). This correlation pattern implies that the ability to make good decisions (in terms of deciding what to do in a situation) is slightly different from the ability to avoid bad decisions. An exploratory factor analysis confirmed this. When these six scores were factor analyzed, a two-factor solution emerged. The two factors were correlated .62.

Table 1. Correlations among SJT Scoring Algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Best</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Worst</td>
<td></td>
<td>.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Best + Worst</td>
<td></td>
<td>.82</td>
<td>.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. 1 – Reverse Best (picked keyed worst as best)</td>
<td>.70</td>
<td>.52</td>
<td>.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. 1 – Reverse Worst (picked keyed best as worst)</td>
<td>.55</td>
<td>.79</td>
<td>.79</td>
<td>.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. 2 – (Reverse Best + Reverse Worst)</td>
<td>.71</td>
<td>.77</td>
<td>.86</td>
<td>.84</td>
<td>.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Best + Worst – Reverse Best – Reverse Worst</td>
<td>.81</td>
<td>.89</td>
<td>.99</td>
<td>.76</td>
<td>.85</td>
<td>.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Key Value of Best</td>
<td>.90</td>
<td>.55</td>
<td>.81</td>
<td>.85</td>
<td>.58</td>
<td>.81</td>
<td>.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. 8 – Key Value of Worst</td>
<td>.51</td>
<td>.94</td>
<td>.87</td>
<td>.52</td>
<td>.89</td>
<td>.83</td>
<td>.89</td>
<td>.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Key Value of Best – Key Value of Worst</td>
<td>.76</td>
<td>.87</td>
<td>.95</td>
<td>.74</td>
<td>.85</td>
<td>.92</td>
<td>.97</td>
<td>.85</td>
<td>.92</td>
<td></td>
</tr>
</tbody>
</table>

Note. Scores are based on all 40 items, n = 1,850. Missing item scores were imputed using the soldier’s mean item score. All correlations are significant at p < .0001.

Table 2 shows the internal consistency reliability estimates and the criterion-related validity estimates of the scoring algorithms. The differences between the algorithms’ validity estimates are small and, in most cases, not statistically significant. These similar validity estimates show that the superior reliability of an algorithm does not necessarily translate into higher validity. For example, algorithm 1 had the lowest
reliability, but its validity estimate is higher (although not significantly higher) than many of the other algorithms.

Algorithm 10 has the highest reliability (tied with algorithm 9) and validity estimates. On a rational basis, it appears to include more information than the other algorithms. It measures both the ability to pick the best action and the ability to avoid the worst action; in addition, it weights the score by the keyed effectiveness value. It is the only scale to include all three of these pieces of information. Therefore, algorithm 10 was used in all subsequent SJT analyses.

The SJT-X used a different scoring algorithm which we called Correctness of Option Rank-Ordering: The absolute difference between the soldier’s ranking of an option and the SMEs’ ranking of the option. Thus, the maximum score is achieved when the soldier puts the options in the same order (in terms of effectiveness) as the SMEs. These absolute differences were summed to produce a total score for the item. The item score was rescaled so that a score of 0 represented random responding and a score of 1 represented a perfect score.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Reliability (coefficient alpha)</th>
<th>Correlation with Observed Performance</th>
<th>Correlation with Future Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Best</td>
<td>.56</td>
<td>.17</td>
<td>.12</td>
</tr>
<tr>
<td>2. Worst</td>
<td>.75</td>
<td>.14</td>
<td>.11</td>
</tr>
<tr>
<td>3. Best + Worst</td>
<td>.78</td>
<td>.18</td>
<td>.14</td>
</tr>
<tr>
<td>4. 1 – Reverse Best (picked keyed worst as best)</td>
<td>.63</td>
<td>.11</td>
<td>.08</td>
</tr>
<tr>
<td>5. 1 – Reverse Worst (picked keyed best as worst)</td>
<td>.71</td>
<td>.14</td>
<td>.11</td>
</tr>
<tr>
<td>6. 2 – (Reverse Best + Reverse Worst)</td>
<td>.75</td>
<td>.15</td>
<td>.12</td>
</tr>
<tr>
<td>8. Key Value of Best</td>
<td>.74</td>
<td>.17</td>
<td>.14</td>
</tr>
<tr>
<td>9. 8 – Key Value of Worst</td>
<td>.84</td>
<td>.15</td>
<td>.13</td>
</tr>
<tr>
<td>10. Key Value of Best – Key Value of Worst</td>
<td>.84</td>
<td>.19</td>
<td>.15</td>
</tr>
</tbody>
</table>

Note: Validity estimates are uncorrected. Scores are based on all 40 items. \( n = 1,567–1,658 \) for the reliability estimates. \( n = 981 \) for observed performance, \( n = 991 \) for future performance. All correlations are significant at \( p < .01 \).

Selection of the Items

Analyses determined that the SJT could be shortened from its original length of 40 items without drastically reducing its quality. We decided, however, to draw an equal number of items from each KSA when constructing the SJT form for the validation. This balanced approach would help to ensure that the test covers a broad range of content.
There is no consensus among test developers about the best method for shortening a test. We compared the criterion-related validity estimates of five different methods. The best 24 items for each method were selected. A double cross-validation design was used to minimize capitalization on chance. These results (see Table 3) suggest that selecting items based on their criterion-related validity estimates leads to the highest cross-validated validity estimate. Picking items based solely on their correlations with the total score (Method 5) yielded the lowest validity estimate (although not significantly lower than most of the other methods). The items for the final E5 and E6 forms were selected based upon the item validities. Separate shortened test forms were created for E5 soldiers and for E6 soldiers. The two 24-item test forms had 12 items in common. Scores for E4 soldiers were computed using the E5 test form.

RESULTS

Descriptive Statistics

The means and standard deviations for the SJT total score are shown in Table 4. Each soldier’s total score was computed two ways: once using the E5 form and once using the E6 form. The mean difference between the E4 and E5 levels was double that between the E5 and E6 levels. This is what one would expect because the amount of training and the number of experiences related to leadership increase much more from E4 to E5 than from E5 to E6. The promotion from E4 to E5 involves profound change—from soldier to NCO. In contrast, promotion from E5 to E6 increases the NCO’s span of control and brings some new experiences, but the types of tasks performed are similar.

Table 3. Criterion-Related Validity Estimates of Different Item-Selection Methods

<table>
<thead>
<tr>
<th>Item Selection Method</th>
<th>Mean Validity Estimate in the 2 Analytic Samples</th>
<th>Mean Validity Estimate in the 2 Validation Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Item-criterion correlations</td>
<td>.240</td>
<td>.197</td>
</tr>
<tr>
<td>2. Correlations with other predictors</td>
<td>.178</td>
<td>.182</td>
</tr>
<tr>
<td>3. Combination of methods 1, 2, &amp; 5</td>
<td>.205</td>
<td>.181</td>
</tr>
<tr>
<td>4. Iterative removal of item with lowest item-scale correlation.</td>
<td>.167</td>
<td>.167 *</td>
</tr>
<tr>
<td>5. Item-total correlations</td>
<td>.150</td>
<td>.150 *</td>
</tr>
</tbody>
</table>

* Cross-validated estimate is significantly lower than the Method 1 cross-validated estimate at \( p < .05 \).

Notes. The observed performance composite was used as the criterion. Methods are listed in descending order of validity estimate in the validation sample. Methods 1, 2, 3, and 5 are based on a 24-item test. Method 4 is based upon 32 items; its validity estimate would likely be lower if based upon a 24-item test. \( n = 485 \) (sample 1) and 486 (sample 2). Correlations are uncorrected. Item selection using Method 4 was done using samples 1 and 2 combined (i.e., it was not cross-validated).
Table 4. Descriptive Statistics by Pay Grade for the Total Score of the SJT

<table>
<thead>
<tr>
<th>Pay Grade</th>
<th>E5 Form</th>
<th>E6 Form</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n$</td>
<td>Mean</td>
</tr>
<tr>
<td>E4</td>
<td>437</td>
<td>1.80</td>
</tr>
<tr>
<td>E5</td>
<td>866</td>
<td>2.19</td>
</tr>
<tr>
<td>E6</td>
<td>545</td>
<td>2.38</td>
</tr>
</tbody>
</table>

**Validity Estimates**

Criterion-related validity estimates were computed using separate forms for E5 and E6 soldiers. These correlations were corrected for criterion unreliability and range restriction. The SJT’s correlations with observed performance ratings were .39 and .25 for E5 and E6 soldiers, respectively. As explained earlier, the 24 items for the final E5 and E6 scores were selected according to their correlations with the two criteria (observed performance composite and expected future performance composite). Therefore, the reported validity estimates are somewhat inflated because the same sample was used to select the items and compute the validities of the total scores. Based on cross-validation results, the shrunken validities are estimated to be .32 and .17 for the E5 and E6 forms, respectively.

The criterion-related validity estimates for the SJT-X were .22 and .18 for future performance and observed performance, respectively. These correlations were corrected for criterion unreliability and for range restriction in the predictor.

**LEADERSHIP STYLES**

The content of the SJT was examined to see whether it could be used to determine which leadership styles a soldier prefers. Six Industrial-Organizational psychologists sorted the response options into piles. Each pile represented actions that the same soldier might be expected to take. This sorting was done independently first. Then the six psychologists discussed their sorting schemes and agreed upon a final set of action categories and the options within each category. These 15 categories appeared to represent leadership styles. The styles are: (a) Do Nothing, (b) Take the Easy Way Out, (c) Shift Responsibility, (d) Delegate, (e) Use Chain of Command, (f) Express Expectations, (g) Preference for Training, (h) Impose Structure, (i) Gather Information, (j) Focus on Tasks Rather than People, (k) Preference for Punishment, (l) Maintain Group Cohesion, (m) Discuss Personal Problems, (n) Express Personal Concern, and (o) Unit Protector or Proponent.

A complex scoring algorithm was developed for these leadership styles. Essentially, each option was associated with one leadership style. A soldier’s score on that style was high if he/she picked many of its options as the best action, low if he/she picked them as the worst action. The score was corrected for the keyed response such that correctly picking the keyed best response or the keyed worst response had less effect on a soldier’s style score than picking another response.
The mean scores on each style were computed separately for E4 soldiers, E5 soldiers, and E6 soldiers. The E5 and E6 means differed very little. E4 soldiers, on the other hand, tended to score higher than E5 soldiers or E6 soldiers on Do Nothing and Unit Protector or Proponent; they tended to score lower on Focus on Tasks Rather Than People, Express Expectations, and Impose Structure. In addition, the mean scores of some of the styles among E4 soldiers, E5 soldiers, and E6 soldiers differed considerably from the mean style scores of the senior NCOs who developed the scoring key. Thus, senior NCOs appear to differ from junior NCOs in terms of which leadership styles are considered most appropriate.

The correlation between each leadership style score and observed job performance rating was computed. These correlations differed considerably between styles. They also differed considerably between E5 soldiers and E6 soldiers. This implies that supervisors (i.e., those giving the job performance ratings) believe that a different balance (i.e., relative importance) of leadership styles should be used for E5 soldiers vs. E6 soldiers. Alternatively, however, it could mean that supervisors of E5 soldiers differ in their opinions (concerning the best balance of leadership styles) from the supervisors of E6 soldiers.

The correlation with performance ratings was higher by at least .05 (i.e., stronger positive correlation or weaker negative correlation) for E6 soldiers compared to E5 soldiers for the following leadership styles: Delegate, Do Nothing, and Shift Responsibility. These correlations for E5 and E6 soldiers were both positive for Delegation and both negative for Do Nothing and Shift Responsibility. For the following styles, the correlation was higher for E5 soldiers: Discuss Personal Problems, Express Expectations, and Gather Information. Thus, it appears that, compared to E6 soldiers, E5 soldiers are expected, by their supervisors, to personally take action more often. It also appears that E5 soldiers are expected to communicate more with people. In particular, E5 soldiers are frequently expected to get more information before making a decision.

**SUMMARY**

The SJT’s high estimates of validity and incremental validity (see Knapp et al., 2002) support its use in helping decide whom to promote to the E5 and E6 levels. These validity estimates are based upon 24-item forms, which would require only about 40 minutes to administer. Separate forms should be developed for the E5 and E6 levels. Knapp et al. show that the effects of race and gender are relatively small. Females and blacks scored almost as high as males and whites, respectively. The differential prediction analyses showed no fairness problems with the SJT.

Although the SJT was developed for promotion purposes, it could also serve as a valuable training tool. The SJT could provide realistic scenarios to E4 and E5 soldiers, which they could use to hone their decision-making skill.

Some pay grade differences emerged in the leadership styles. E4 soldiers’ style preferences differ somewhat from those of E5 soldiers and E6 soldiers. Supervisors, on
the other hand, seem to think that E5 soldiers and E6 soldiers *should* differ somewhat in their leadership style. Further research is needed to examine these and related issues more closely. These leadership style scores have many potential uses such as test development, training, performance measurement, and leadership research.

The SJT-X targets a relatively narrow construct: *Knowledge of the Inter-Relatedness of Units*. The SJT-X had respectable criterion-related validity estimates in spite of imperfect criteria. Although most supervisors provided performance ratings for a construct similar to this one (*Coordination of Multiple Units and Battlefield Functions*), it is unlikely that many supervisors have actually observed their subordinates in situations relevant to this construct. Thus, it is possible that the validity estimates would have been higher had a better criterion measure been available.

**REFERENCE**

PAY GRADE DIFFERENCES IN THE FUNCTIONING OF THE SITUATIONAL JUDGMENT TEST  

Dan J. Putka, Ph.D.  
Research Scientist  
Human Resources Research Organization  
66 Canal Center Plaza, Alexandria, VA 22314  
dputka@humrro.org  

Gordon W. Waugh, Ph.D.  
Senior Scientist  
Human Resources Research Organization  
66 Canal Center Plaza, Alexandria, VA 22314  
gwaugh@humrro.org  

Deirdre J. Knapp, Ph.D.  
Manager Assessment Research and Analysis Program  
Human Resources Research Organization  
66 Canal Center Plaza, Alexandria, VA 22314  
dknapp@humrro.org  

INTRODUCTION  

In its initial development, a single, 40-item version of the NCO21 Situational Judgment Test (SJT) was created for use in promoting Corporals and Specialists (E4 soldiers) to Sergeants (E5 soldiers) and Sergeants to Staff Sergeants (E6 soldiers) in the U.S. Army. In the process of concurrently validating this instrument against supervisors’ ratings of soldiers’ observed performance, substantial differences in the validity estimates were found depending on whether the participating soldier was an E5 or E6 soldier (Sager, Putka, & Burnfield, 2002). These differences in validity estimates remained even after correcting for common statistical artifacts that may have been differentially affecting the validity estimates at each pay grade (e.g., range restriction, criterion unreliability). In the present paper, we evaluated two plausible explanations for differences in SJT validity estimates. First, the SJT may be tapping different classes of predictor constructs (e.g., cognitive aptitude, experience, and temperament) for E5 and E6 soldiers. Second, the validity estimate of the SJT among E6 soldiers may be attenuated by the inclusion of highly experienced Staff Sergeants in the concurrent validation sample. In the present paper, we further investigate the tenability of these explanations for differences in SJT validity estimates across pay grades.

---

5 This paper is part of a symposium titled Tailoring a Situational Judgment Test to Different Pay Grades presented at the 2002 International Military Testing Association Conference in Ottawa (G.W. Waugh, Chair). The views, opinions, and/or findings contained in this paper are those of the authors and should not be construed as an official U.S. Department of the Army position, policy, or decision.
The SJT Measures Different Constructs at Different Grades

Past research has indicated that cognitive aptitude, experience, and temperament are all related to SJT scores (McDaniel, Morgeson, Finneghan, Campion, & Braverman, 2001). Differences in the strength of these relationships across pay grades may help determine why criterion-related validity estimates for the SJT vary across grades. For example, the criterion used for validating the SJT in the NCO21 validation effort was supervisor ratings. Such ratings often reflect “will-do” or motivational aspects of performance, which tend to be more closely related to temperament constructs than to cognitive aptitude (Borman & Motowidlo, 1997). Thus, higher validity estimates for the SJT among E5 soldiers might result if the SJT is a more a function of temperament for Sergeants than for Staff Sergeants.

While such findings would help explain differences in SJT validity estimates among E5 and E6 soldiers, they would not address why the SJT might assess different constructs at different pay grades. For example: Considering that the same SJT was administered to all soldiers under the same instruction set, how can the SJT be assessing different constructs at different pay grades? An examination of the characteristics of the items that comprise the NCO21 SJT might answer this question.

The majority of items on the NCO21 SJT put respondents into a situation in which they are playing a given role (e.g., “you are a Sergeant”, “you are a Corporal”). Some of these roles are based at the E4 soldier level (7 of 40 items), some at the E5 soldier level (14 of 40 items), and some at the E6 soldier level (3 of 40 items). Given that the SJT was designed to differentiate between promotion-eligible E4 soldiers, and between promotion-eligible E5 soldiers, it is not surprising that most of these role-specific items ask respondents to assume the role of an E4 or E5 soldier (21 of 24 items). Depending on the pay grade of the respondent completing the SJT, as well as the role to be played in the given item, the cognitive task of choosing the most and least effective options may be quite different, thus creating the potential for different constructs to be assessed at each pay grade.

For example, when E5 soldiers respond to items in which they are asked to play the role of a Sergeant, there is no inferential leap required for them to assume a role at a different pay grade. The question for E5 soldiers on such items is one of what they should do in that situation. When E6 soldiers answer the same item, the question is slightly different: They must put themselves in the shoes of an E5 soldier. It is important to note that the instruction set given to E5 and E6 soldiers is the same for all items. Assimilating a different role may add to the cognitive complexity of completing the SJT. Given the current structure of the SJT, E6 soldiers must assume a different role much more frequently than E5 soldiers when completing the test. Thus, cognitive aptitude might affect SJT scores more among E6 soldiers because of the greater cognitive complexity of the SJT for E6 soldiers, relative to E5 soldiers. Furthermore, the cognitive complexity of the SJT might be greatest for E6 soldiers who have not been an E5 soldier for a long time.
The SJT was Not Designed for Use Among E6 Soldiers

Based upon this reasoning, we explored another possible explanation for pay grade differences in the relationship between SJT scores and performance. We examined whether the relationship became weaker as the amount of time as an E6 soldier increased. One characteristic of promotion in the Army is that as soldiers advance to upper pay grades, the time between promotions to subsequent grades becomes longer. Indeed, there is likely to be a much greater range of leadership experience among E6 soldiers compared to E5 soldiers. As stated earlier, the SJT was designed for use in E4-to-E5 and E5-to-E6 soldier promotions. Inclusion of high-experience E6 soldiers in the validity analyses might make observed validity estimates misleadingly low when the SJT is used to promote eligible E4 and E5 soldiers.

The Present Paper

The present paper examined the potential explanations discussed above. Specifically, we set out to answer the following questions:

?? Does the strength of the relationship between SJT scores and cognitive aptitude, experience, and temperament differ between pay grades?

?? Are SJT scores more strongly related to cognitive aptitude for E6 soldiers than for E5 soldiers?

?? Is the relationship between cognitive aptitude and SJT scores among E6 soldiers moderated by time in pay grade? Specifically, does cognitive aptitude positively predict SJT performance better for high experience E6 soldiers than for low experience E6 soldiers?

?? Is the relationship between observed performance and SJT scores moderated by time in pay grade among E6 soldiers? Specifically, do SJT scores positively predict performance better for low experience E6 soldiers than for high experience E6 soldiers?

METHOD

Sample

Data were drawn from the NCO21 validation database. These data were collected from seven different Army bases between April and August of 2001. The sample examined in the present paper was based on data from 866 E5 soldiers and 545 E6 soldiers.
Measures

NCO21 SJT

The development and validation of the SJT has been extensively described elsewhere (Waugh, Putka, & Sager, 2002). For E6 soldiers the mean SJT score was 2.35 ($SD = 0.48$); for E5 soldiers it was 2.16 ($SD = 0.55$).

Cognitive Aptitude

The ASVAB General Technical (GT) score is a composite of three subtests from the Armed Services Vocational Aptitude Battery (Word Knowledge, Paragraph Comprehension, and Arithmetic Reasoning). The ASVAB provides the basis for selecting and classifying new recruits in all of the U.S. Armed Services. The GT score is an indicator of general cognitive aptitude.

Experience

The Experiences and Activities Record (ExAct) has 46 self-report items assessing past experience in a variety of activities. Two scores derived from the ExAct in the NCO21 validation effort were used in the current investigation: Supervisory Experience and General Experience.

Temperament

The Assessment of Individual Motivation (AIM) is a forced-choice measure that provides scores for six temperament constructs (i.e., Dependability, Adjustment, Work Orientation, Agreeableness, Physical Conditioning, and Leadership; Young, Heggestad, Rumsey, & White, 2000). The BIQ is a self-report measure consisting of biodata-like items that provide scores on eight constructs relevant to future NCO performance (i.e., Hostility to Authority, Manipulativeness, Social Perceptiveness, Social Maturity, Tolerance for Ambiguity, Openness, Leadership, and Interpersonal Skill; Kilcullen, Mael, Goodwin, & Zazanis, 1999).

Observed Performance Ratings

The Observed Performance Rating (OPR) Scales ask soldiers’ supervisors to rate their subordinates using 19 specific performance scales, where each scale taps a critical NCO performance requirement. These scales were developed as part of the NCO21 research effort (Knapp et al., 2002). The mean of 18 of these scales was used as a performance criterion for each soldier. Interrater reliability estimates, as indicated by $ICC(C,k)$ for this composite were .53 for E5 soldiers and .59 for E6 soldiers (McGraw & Wong, 1996).
RESULTS

Table 1 shows the raw zero-order correlations between the SJT and cognitive aptitude, temperament, and experience-based predictors. It also shows standardized beta weights, as well as the relative weights associated with these predictors when they were all used to predict SJT scores simultaneously (Johnson, 2001).
Table 1. Relationships between the Predictor Measures and SJT Scores by Pay Grade

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Raw Correlation with SJT</th>
<th>Standardized Beta Weight</th>
<th>Relative Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E5</td>
<td>E6</td>
<td>E5</td>
</tr>
<tr>
<td>Cognitive Aptitude</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASVAB GT</td>
<td>.15</td>
<td>.25</td>
<td>0.10</td>
</tr>
<tr>
<td>Experience</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ExAct: Supervisory Experience</td>
<td>-.05</td>
<td>-.02</td>
<td>-0.06</td>
</tr>
<tr>
<td>ExAct: General Experience</td>
<td>.05</td>
<td>.01</td>
<td>0.04</td>
</tr>
<tr>
<td>Temperament</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIM: Dependability</td>
<td>0.36</td>
<td>0.22</td>
<td>0.16</td>
</tr>
<tr>
<td>AIM: Adjustment</td>
<td>0.24</td>
<td>0.14</td>
<td>0.00</td>
</tr>
<tr>
<td>AIM: Work Orientation</td>
<td>0.28</td>
<td>0.13</td>
<td>0.02</td>
</tr>
<tr>
<td>AIM: Agreeableness</td>
<td>0.31</td>
<td>0.21</td>
<td>0.11</td>
</tr>
<tr>
<td>AIM: Physical Conditioning</td>
<td>0.12</td>
<td>0.10</td>
<td>0.00</td>
</tr>
<tr>
<td>AIM: Leadership</td>
<td>0.28</td>
<td>0.20</td>
<td>0.21</td>
</tr>
<tr>
<td>BIQ: Hostility to Authority</td>
<td>-.33</td>
<td>-.20</td>
<td>-0.12</td>
</tr>
<tr>
<td>BIQ: Manipulativeness</td>
<td>-.34</td>
<td>-.19</td>
<td>-0.06</td>
</tr>
<tr>
<td>BIQ: Social Perceptiveness</td>
<td>0.13</td>
<td>0.01</td>
<td>0.05</td>
</tr>
<tr>
<td>BIQ: Social Maturity</td>
<td>0.27</td>
<td>0.20</td>
<td>0.01</td>
</tr>
<tr>
<td>BIQ: Tolerance for Ambiguity</td>
<td>0.20</td>
<td>0.10</td>
<td>-0.01</td>
</tr>
<tr>
<td>BIQ: Openness</td>
<td>0.07</td>
<td>-0.01</td>
<td>0.05</td>
</tr>
<tr>
<td>BIQ: Leadership</td>
<td>0.12</td>
<td>0.11</td>
<td>-0.11</td>
</tr>
<tr>
<td>BIQ: Interpersonal Skill</td>
<td>0.33</td>
<td>0.18</td>
<td>0.06</td>
</tr>
<tr>
<td>Multiple R</td>
<td>0.50</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>Shrunken Multiple R</td>
<td>0.47</td>
<td>0.33</td>
<td></td>
</tr>
</tbody>
</table>

Notes. \(n_{E5} = 800\) (listwise); \(n_{E6} = 480\) (listwise). Relative weights reflect the percentage of \(R^2\) accounted for by each predictor. Italicized relative weight values for experience and temperament are the sum of the relative weights for the predictors within that class. Statistically significant correlations and beta weights are in boldface \((p < .05)\).

The pattern of correlations and regression-based weights illustrate consistent differences between E5 and E6 soldiers. As hypothesized, cognitive aptitude appears to account for more variation in SJT scores for E6 soldiers (Relative Weight = 31.6%) than for E5 soldiers (Relative Weight = 5.3%). On the other hand, temperament measures appear to account for more variation in SJT scores for E5 soldiers (Sum of Relative Weights for Temperament = 92.3%) than for E6 soldiers (Sum of Relative Weights for Temperament = 67.6%), as evidenced by the consistently higher correlations among E5...
soldiers. Interestingly, experience-based measures appeared to have very little relationship with SJT scores among either E5 or E6 soldiers.

Next, we examined whether time in grade moderated the relationship between cognitive aptitude and SJT scores. A moderated multiple regression analysis revealed that, when predicting SJT scores among E6s, there was no significant interaction between time in grade and cognitive aptitude. Thus, no evidence emerged to suggest that the relationship between cognitive aptitude and SJT scores was stronger for E6s who have more time in grade.

Finally, we addressed the possibility that time in grade acts as a moderator of the relationship between SJT scores and observed performance among E6 soldiers. Moderated multiple regression analyses revealed significant main effects for time in grade ($\beta = 0.75$, $p < .01$) and SJT ($\beta = 0.23$, $p < .01$), as well as their interaction ($\beta = -0.70$, $p < .01$). Figure 1 illustrates the nature of this interaction.

![Figure 1. Moderating Effect of Time in Grade on the Relationship between SJT Scores and Observed Performance for E6 Soldiers](image)

Two regression lines are plotted in Figure 1, one where time in grade is set equal to $+1$ standard deviation (high time in grade), and the other where time in grade is set equal to $-1$ standard deviation (low time in grade). As Figure 1 illustrates, there is a disordinal interaction between time in grade and SJT scores when predicting observed performance. Specifically, among E6 soldiers with low time in grade, SJT scores are strongly positively related to observed performance; in contrast, among E6 soldiers with high time in grade, the relationship is negative. This finding provides support for the claim that the SJT validity estimates for E6 soldiers are lower than the validity estimates for E5 soldiers because high-experience Staff Sergeants are included in the NCO21 validation sample.
DISCUSSION

In the course of validating the NCO21 SJT against an observed performance criterion, differences were found in the validity estimates for E5 and E6 soldiers. This paper investigated two potential explanations for these differences. First, we found that SJT scores tended to be more a function of cognitive aptitude for E6 soldiers compared to E5 soldiers, though temperament-related predictors accounted for substantial amounts of variance in SJT scores for both E5 and E6 soldiers. Given that “will-do” criteria such as supervisor ratings tend to covary more with temperament, rather than cognitive aptitude, these findings provide one explanation for lower validity estimates among E6 soldiers. Second, we found that E6 soldiers’ time in grade moderated the relationship between SJT scores and observed performance. The relationship between SJT scores and observed performance for E6 soldiers with low levels of time in grade was strongly positive, whereas the relationship between SJT scores and observed performance for E6 soldiers with high levels of time in grade was negative. This pattern of findings suggests that inclusion of high tenure Staff Sergeants in the original NCO21 validation sample attenuated E6 soldier validity estimates relative to E5 soldier estimates.

The implications for future development and validation of SJT are relatively straightforward. When developing SJTs that will be used to promote personnel at varying levels of experience, rank, or pay grade, special attention should be paid to balancing items in terms of the roles respondents are asked to take on in each of the items. Doing so will help ensure that respondents at all levels are required to assimilate roles different from their own position to similar degrees, thus better equalizing the cognitive complexity of the testing task for all respondents. Second, when conducting concurrent validation studies for promotion tests, one must carefully consider the composition of the validation sample relative to the relevant population, being particularly wary of differences in experience and pay grade. Third, it might help examinees to assume the appropriate roles if the items in an SJT are grouped by role. For example, all of the E4 soldier items could be listed under the heading, “In the next set of questions, assume the role of an E4 soldier.” It might also help to explicitly state, in every test item, the pay grade the examinee should assume. Finally, for many SJTs, it might be possible to avoid the problem by either writing all items to the same pay grade (e.g., targeting E5 soldiers) or omitting pay grade information entirely (e.g., scenario descriptions could use the terms, superior, subordinate, and peer).

REFERENCES


A MODEL FOR THE CONCURRENT DESIGN OF PERFORMANCE

Robert K. Branson
Director, Center for Performance Technology, Florida State University, Tallahassee, FL
32306-2540
rbranson@cpt.fsu.edu

INTRODUCTION

One arena that has a vivid current need for new conceptualizations is that of training and human performance development, often called the "Systems Approach to Training," (SAT) or "Instructional Systems Development," (ISD). This arena began to thrive during the 1960s and subsequently became widely used in the military and in corporate programs. Early ISD/SAT models specified a predominant flow and explicit order of activity. Models published by Gagné, Briggs, and Wager (1974), Dick and Carey (1978), Branson, Rayner, Cox, Furman, King, & Hannum (1975), and the military services reflect these earlier views.

The Model for the Concurrent Design of Performance (CDP) strives to address the traditional and developing needs environments to support individually managed careers in the profession of arms from day one to day N. The CDP Model incorporates the past 30 years of research in performance technology, remote delivery of information and instruction, change management, and relies more heavily on individual initiative than did traditional structures.

A SYSTEMIC VIEW

One significant purpose of the CDP Model is to schedule critical attributes of the total work process in an efficient manner. Figure 1 presents a top-level block diagram to describe the various functions essential to implementing and sustaining effective new processes. The concurrent design model has four distinguishing characteristics:

?? It introduces the process of Concurrent Feedback and Interaction through the "Red Team." The "Red Team" is an invited group of professionally competent critics who are, preferably, not otherwise involved in the program.

?? It is non-linear in the sense that all of the major processes must be conducted simultaneously from the beginning.

?? The model provides for total stakeholder involvement throughout the life cycle.

?? To be consistent with current philosophy and doctrine in military and civilian organizations, the acquisition responsibility for individual and team performance has been delegated to the individual.

The most significant departure from traditional ISD/SAT models comes from the industry practice of concurrent design and engineering, in which cross-functional teams...
are charged with developing new processes and products (Watson, 1993). All legitimate stakeholders are involved from the outset and all must be heard. Consensus of the stakeholders is the predominant decision strategy. Among others, these stakeholders comprise doctrine and strategy developers, analysts, designers, developers, implementers, controllers, and end users or customers.

**Figure 1. A model for the concurrent design of performance.**

The CDP Model synthesizes current understanding of processes from manufacturing, service, engineering, and organizational design, among others. No one function in the model is static or linear; each depends upon the one before, after, and across from itself.

**SUBSYSTEM ONE: ANALYZE CONTEXT AND STAKEHOLDERS**

This group’s function establishes a clear relationship between current operations, the mission, the environment, the threats, and the stakeholders. Here stakeholders comprise all people and organizations that can have either a positive or negative influence on plans or results. The world remembers numerous change advocates who discovered strong opposing stakeholders far too late to avoid defeat (Rogers, 1995).

The outputs from Subsystem One include all subordinate performance goals. The descriptive documents and slideshows must include sufficient detail so that other process owners and affected units will be able to contribute to and critique them.

**CONCURRENT FEEDBACK AND INTERACTION: THE RED TEAM**

The center function in Figure 1 is titled "Concurrent Feedback and Interaction." The "Red Team" is an invited, diverse group of professionally competent critics who are not otherwise involved in the program. The project leaders must find sophisticated...
professionals, representative of all stakeholders, who understand the problem and the solution alternatives, and are capable of evaluating the likelihood of success. The Red Team is a surrogate quality assurance function inserted early in program development. The Red Team must decide whether the solution, if perfectly executed, would solve the problem by using simulated results or projections. Because specifications and measures remain to be developed, the Red Team must rely on effective analytical processes, prior experience, qualitative methods, and consensus.

To begin the Concurrent Feedback and Interaction function, the program managers present the complete performance development concept in a briefing session to the Red Team, who must meet the challenge presented by Subsystem One. They must answer the fundamental question about whether the problem as defined can be solved. Further, if the problem is solved as defined, will that solution have the intended effect on the selected outcome variables?

SUBSYSTEM TWO: DEFINE PERFORMANCE REQUIREMENTS

Subsystem Two uses the outputs from Subsystem One and the Red Team to begin specifying products and processes to achieve the mission. It transforms the mission and goals into design and performance requirements. In a sense, Subsystem Two defines organizational and mission requirements by explicitly detailing issues and feeding back problems, requirements, and questions to Subsystem One. The feedback loop always includes the question: “How well do these requirements match your vision, mission, and goals?”

Requirements may refer to physical space, machine capabilities, performance capabilities, and other defining features that enable systems designers in Subsystem Four to create potential solutions. “Performance requirements” refers to that set of attributes of a system that state in specific and measurable terms what the operating results should be. Following is one example of an outside-in performance requirement:

\[
\text{The average probability of a first round hit by an M-1 tank gunner at 3000m is now } p = .35. \text{ To win, we must raise that average to } p = .65. \text{ Provide the capability to achieve that goal.}
\]

How one meets this performance requirement will depend substantially on the characteristics of the organization and the range of possible solution strategies within the sphere of influence of the program manager.

Sometimes the culture and politics of the organization make the work of the Red Team exceptionally difficult. Several years ago, we were asked to conduct a job analysis of a law enforcement agency and to recommend the content of a training program based on the analysis. In the job analysis, we found no instances in which the officer was expected to write Spanish. We recommended to the sponsors that they could save a great deal of time and money if they removed the writing portion from the course because that requirement was not found in the field. Unfortunately, the course instructors were such a powerful block that they were able to require writing instruction. In this case, the Red Team failed to prevent a bad decision.
The Red Team must decide whether the design requirements can be realistically achieved within the cost and time parameters specified. Further, they must decide whether the requirements have fully addressed the need defined in Subsystem One.

SUBSYSTEM THREE

In Subsystem Three, all major process and outcome measures are developed. The aggregate of these measures must reflect both operations and results. Process and outcome measures comprise major areas of contention in many human performance systems. Gaining consensus on the variables that will be used to evaluate internal and external results often requires a substantial investment in time from the program managers and key stakeholders.

Arguments about these variables emerge from different vested interests. Regardless, major systemic success cannot logically be achieved if the key stakeholders cannot reach consensus on measures of the critical success factors. The reason is simple: If successful outcomes cannot be defined in advance, they cannot be measured. With no consensus measures, success cannot be achieved because the measures present the only credible evidence of success.

The interaction between Subsystems Two and Three provides a strong case for the argument that the concurrent model is non-linear. The requirements definition and results measurement processes cannot be comfortably separated because it is the measurement activity that defines what is really meant by the requirement. The output of Subsystem Three is the list of measures and characteristics that will assess the critical success factors. Elaboration through slideshows will almost certainly be required.

SUBSYSTEM FOUR

Subsystem Four uses inputs from all other subsystems to conduct the systematic design process in which various alternatives are considered and thorough cost-benefit tradeoff analyses are conducted. Based on the constant feedback from the Red Team and interaction with Subsystems One, Two, and Three, solution designers reduce requirements to processes and create the rapid prototype. In concurrent design, the prototyping process is done in three phases:

Phase I is concept or solution validation. The design concept is flow-charted, diagrammed, drawn, and described to the stakeholder community, the research and technology community, and the mission and requirements definition community as a means of validating the high-level concept rendering.

Phase II: Although different occupations and professions have their way of rendering and presenting design, in the field of human performance technology perhaps the most cost effective methodology can be adapted from the film and video industries in the form of a treatment, script, and storyboard.
Phase III: Solution designers can select among the elaborated priorities to design a system that will be just good enough for initial operations. This process moves the models almost seamlessly to the acceptance, implementation, and operations subsystem.

IMPLEMENTATION

When implementation begins, the high payoff for adopting the CDP will begin to appear. We recognize two kinds of implementation processes: social marketing or diffusion and mandated implementation. As numerous system developers have discovered over the years, the implementation process is fraught with unanticipated obstacles. In this phase, the targets are expected to accept the process changes or adopt the innovation as designed. Remember that under a diffusion model, the targets have the choice to adopt or not adopt the program. Such models are found in public education, diffuse social systems, most industries, and many military commands.

Mandates are either hard or soft. In a hard mandate, the commander can use the carrot and stick to obtain desired compliance. In a soft mandate, the authority may only recommend by listing the advantages and disadvantages. Your physician can urge you to adopt an exercise program and your commander can require it. Although military commands do not have a monopoly on failed mandates, they certainly have an impressive market share. The four-star in command A does not mandate processes to the four-star in command B. Mandated changes increase the mental workload of the incumbents who already believe they have a full desk and they will often not put the necessary changes in proper priority.

SUBSYSTEM FIVE: CONTROL

To many human performance technologists (HPTs), the term control often has a negative connotation. For many academic writers and other designers, "control" refers to top-down micromanagement, but this idea is far distant from the intention of a system control function. To avoid the connotation of top-down micromanagement, the academic community that works in HPT changed the system construct of control to "evaluation." One of the reasons that I believe they made that unfortunate substitution is that evaluation is the last event that occurs before write-up and sending the article to the journal.

When the whole system is functioning, either at the beta test or during the early stages of the 1.0 release, the Red Team reviews results, standards, and processes. Suppose that our performance intervention was a training program designed to improve tank gunnery first round hits. The chosen process was that of practice using a laser simulator. The output of Subsystem Five is continuous and detailed. That subsystem reports on all process and outcome measures and compares their status with the expected results.

SUMMARY AND CONCLUSIONS

In this paper, I have attempted to reconcile the processes typically used to design and implement performance and to identify potential areas in which prior approaches
may have been unsuccessful. Two major causes of failure were examined, concept errors and implementation errors. To overcome the all-too-frequent tendency toward concept error, typically found in situations where problems are incompletely analyzed and horizons ineffectively scanned, I have proposed a process redesign—concurrent design and engineering—as a specific means to improve performance.

Many expensive performance improvement failures could have been prevented had the project leaders had the “whole system in the room” from the outset. Many such programs are designed and developed at a distance from those who must implement. The end-users can provide valuable input about potential sources of error. Those advocates who urge the program changes are highly optimistic about its potential success. To move toward a more objective analysis of proposed solutions, I have advanced the concept of the Red Team. The Red Team is a group of professionals not included in the program design who are charged with critique and oversight of the entire program life cycle.

Finally, a concurrent design model represents an attempt to integrate recent research and other valid knowledge into the performance improvement process as a means of guiding programs to successful implementation.

References


The Selection of Military Police: 
An Assessment Centre Approach

Conference Paper 2002-03

Captain J. Klammer, MSc
and
L. Tanner, BScH

September 2002

Reviewed by: 
R. Boswell, MA
Lieutenant-Colonel
Research Co-ordinator

Approved by: 
C.D. Lamerson, Ph.D.
Colonel
Director

44th Annual Conference of the International Military Testing Association
Ottawa, Canada, 22-24 October 2002
INTRODUCTION

Background

The Military Police Assessment Centre (MPAC) was established in 1999 for the purpose of selecting candidates for the MP occupation. Based primarily on selection practices used for the Canadian military, this procedure involved a series of measures that were designed to assess the candidate’s personality attributes and suitability for employment in the MP occupation. These included a file review, a written communications skills test, a personality measure (scored, but not used in the final selection decision), a group dynamics exercise (not scored), and a structured interview.

The original MPAC evolved from a rigorous selection program that had been developed between 1996 and 1999 for CF members applying through occupation transfer programs to the MP occupation (O’Brien, 1996). As of 1998, it was also used to select individuals for component transfers (transferring from the Reserve Force to the Regular Force) and the Air Reserve Augmentation Force (Woycheshin, 1998). The in-service processing procedure was refined between 1999 and 2000 for use with civilian applicants to the CF (Noonan, 1999, 2000) and was further revised in 2001 for the purpose of updating and improving the selection process.

Given the recent re-structuring of the MP Branch to emphasize the importance and responsibility of police work, several civilian police forces across Canada were consulted in 2002 to examine their “best practices” for selecting police officers. These police forces use a variety of measures, including aptitude tests, physical fitness tests, written communication skills tests, video simulations, role playing, group dynamics exercises, background/integrity interviews, structured panel interviews, and polygraph examinations. Based on these consultations, several new measures were incorporated into the MPAC for 2002.

DESIGNING THE MPAC

Assessment Centre Method

In the assessment centre method, candidates are engaged in a variety of job-related exercises designed to simulate realistic situations that an individual would face when working in a particular job. These exercises are designed to elicit behaviour relevant to the abilities, skills, and personal qualities that are critical to success in the targeted job. Assessors who are familiar with the target position and who have been
trained in the assessment centre method observe and evaluate the candidate’s behaviour. The key to any successful assessment centre is to use multiple tools and multiple observers to assess each candidate.

The MPAC research team decided to formalize its assessment approach by following the guidelines and ethical considerations for assessment centre operations found in the ICACM model (ICACM, 2000). This was seen as important to ensure:

a) Legal defensibility of the system,
b) Inclusion of a number of selection activities with high face validity,
c) A longer training session for assessors, and
d) Exposure of the candidates to a wide range of realistic MP activities.

**MP Competencies**

Concurrent to the consultations with civilian police forces, the list of competencies originally developed for the MP occupation in 1996 was revised. This revision was necessary in light of the availability of additional documentation and recent changes in the MP occupation. An occupational analysis of the MP occupation conducted in the late 1990s and the original job analysis conducted in 1996, both of which had used subject matter experts in gathering the data, were reviewed. Information gathered from these analyses, as well as input from currently serving members of the MP Branch solicited through a survey, was the basis of the revised set of 11 competencies, shown in Table 1.

**MPAC Design**

The MPAC is a two and a half-day program designed to assess a maximum of 36 candidates per serial. There are often numerous serials run back-to-back during each MPAC, with approximately two MPACs held each year. The MPAC is highly structured with the following features:

a) Three-day training session for assessors prior to each MPAC;
b) Observer rating forms, assessment guides and competency anchors;
c) Standardized exercise openings, instructions, and closings; and
d) Continuous observation of candidates.
### Table 1

**MP Competencies**

| **Integrity** | – Adheres to the values of honesty and trustworthiness.  
– Resists temptations of an unethical or illegal nature. |
| **Leadership** | – Takes charge; establishes priorities; provides team with a clear sense of direction and purpose.  
– Accepts responsibilities for outcomes.  
– Facilitates successful goal accomplishment by inspiring/motivating others, by example, to perform to the highest standards.  
– Keeps others informed of issues/concerns; advises on courses of action and potential consequences.  
– Delegates responsibilities to appropriate subordinates; gives others latitude to exercise their own initiative, and invests them with the power and authority to accomplish tasks effectively.  
– Provides counselling as required on career and leadership issues. |
| **Management Control** | – Produces both short and long term plans that are detailed, with clearly defined objectives.  
– Coordinates and monitors the execution of tasks/projects; schedules activities to ensure optimum use of time and resources (human, material, and financial) so that key objectives are met. |
| **Analytical Skills** | – Seeks all possible relevant information for problem solving and decision making; consults widely, probes the facts carefully, and thoroughly and critically analyzes issues from different perspectives.  
– Does simple math and numerical calculations in a business context. |
| **Decision Making** | – Makes rational, realistic and sound decisions based on consideration of all available facts and alternatives.  
– Makes firm and timely decisions and commits to definite courses of action, on the basis of limited information if necessary. |
| **Personal Impact** | – Projects a good first impression, commands attention and respect, shows an air of confidence and achieves personal recognition.  
– Confident in own abilities and judgement, while understanding own limitations. |
| **Interpersonal Skills/Tolerance** | – Effectively and respectfully interacts with individuals of different backgrounds, personalities, attitudes, opinions and values.  
– Shows sensitivity, compassion and sincerity.  
– Tactful, yet diplomatic.  
– Uses attending skills when interacting with others (e.g., establishes eye contact, paraphrases, demonstrates interest, etc.). |
| **Conscientiousness** | – Efficient, thorough, hardworking and dependable.  
– Meets or exceeds given standards and deadlines.  
– Persistently strives for excellence even in difficult situations. |
| **Stress Tolerance/Self-Control** | – Keeps own emotions under control and restrains from negative actions when provoked or when working under stressful conditions.  
– Effectively manages stress to prevent it from negatively impacting performance. |
Table 1

MP Competencies (cont)

| Initiative          | – Is self-motivated and self-directed in identifying and addressing important issues.  
<table>
<thead>
<tr>
<th></th>
<th>– Actively influences events rather than passively responds to them.</th>
</tr>
</thead>
</table>
| Flexibility         | – Adapts one’s approach to a variety of situations.  
|                     | – Receptive to innovation; accepts the non-traditional or embraces change.  
|                     | – Generates new and imaginative ideas; creates novel solutions. |
| Practical Intelligence | – Possesses good judgement and common sense.  
|                     | – Recognizes and uses changing information; evaluates and utilizes various courses of action.  
|                     | – Not overly reliant on logic; reacts instinctively in emergencies.  
|                     | – Understands the dynamics of organizations, including the formal and informal cultures and decision-making processes. |
| Oral Communication Skills | – Clearly, accurately and concisely expresses and listens to ideas, feelings, questions and facts in both individual and group situations. |
| Written Communication Skills | – Accurately and concisely describes events and presents conclusions, organizing written material in a clear, logical manner using appropriate grammar, style and language. |

**MPAC Components**

The MPAC has four main components – assessor training, candidate assessments, orientation, and the selection board. Assessor training is essential to ensure good inter-rater reliability; the assessment phase evaluates candidates (from the results of several MP-relevant measures) in terms of their potential for successful performance as an entry-level MP; the orientation component is aimed at providing MP applicants with realistic expectations regarding a career as an MP; and the selection board examines the collated information on each candidate and makes a selection recommendation to the Canadian Forces Provost Marshal (CFPM).

**MPAC Training.** For the MPAC to be useful as a selection device, assessors are trained to understand and recognize the competencies important to MP job performance, observe and record the behaviour of candidates during the assessment exercises, categorize the candidates’ behaviour according to the appropriate competencies, and use this data to rate the candidates on each dimension. In effect, this training “calibrates” the assessors’ scores and ensures that there is good inter-rater reliability. If assessors are not adequately trained in observing and rating methods, the value of their evaluations will be lessened. In addition, in keeping with current legislation, training is given on the Canadian Human Rights Act to ensure proper guidelines are followed.

**Assessment Phase.** Each assessment team monitors a group of 10-12 candidates. The assessment team is composed of six individuals, one of whom is the team leader. It is the team leader’s responsibility to ensure consistency in the evaluation of the
candidates and the completion of forms required for the Selection Board. Of the remaining five assessors, one is normally a Personnel Selection Officer (PSO) or another individual from outside the MP Branch (e.g., a Military Career Counselor), which ensures the input of an independent observer, while the remainder is comprised of senior MPs and Military Police Officers (MPOs).

Structured assessments begin the night of the candidates’ arrival, and continue through the next day. During this time, assessment team members record observations of behaviour continuously throughout a series of activities. A strict schedule is followed to ensure that an assessor does not assess a candidate more than once. Once an assessment exercise has been completed, assessors document and score the candidate they have observed. Scores are then compiled with those provided by other assessment team members, each of who will have observed the same candidate in a different exercise. As a summary of the candidate, each assessor’s scores are entered in a composite score report for presentation during the Selection Board. A senior MPO is appointed by the CFPM to chair the MPAC process and the final MPAC Selection Board. To maintain impartiality, the Chair does not form part of an assessment team.

The MPAC requires candidates to complete all exercises. Candidates who refuse to attempt an exercise or who are unable to continue for whatever reason are removed from the activity. Failure to complete an exercise does not automatically result in “failure” per se, but it does negatively impact the selection decision. Candidates who remove themselves from the assessment process or who are not selected are normally allowed to re-attempt at a later date.

Orientation Phase. In addition to helping ensure candidates forge realistic career expectations through receiving information on such matters as MP training, job duties, career patterns, and the military lifestyle, the orientation program is a concrete expression of the military’s interest in the candidate. Numerous organizations have found that realistic orientation programs increase commitment to the organization and reduce attrition (Meglino, DeNisi, Youngblood & Williams, 1988; Wanous, Poland, Premack & Davis, 1992). The orientation phase includes a realistic job preview (RJP), which features the following components:

a) senior MPO/MP briefings on the MP role in the CF and training progression;

b) briefings/visual recruiting products on specific aspects of MP duties, with a particular focus on law enforcement duties in domestic and international operations; and

c) discussions on career/lifestyle aspects, stressors and demands.

Where feasible, the orientation phase may include additional RJP activities such as tours of military and MP facilities and demonstrations of MP equipment. These are conducted after the assessment phase has been completed to ensure fairness to all candidates. It is possible that candidates receiving familiarization visits prior to certain assessment activities would have an unfair advantage in that they could better visualize
some of the scenarios presented to them. For this reason, familiarization visits are not conducted until the assessment phase is over.

**Selection Board.** Each Assessment Team Leader is responsible for presenting, to the Selection Board Chair, the composite score reports for each candidate the team assessed. Included in the report are notes on the candidate’s evaluations. For example, where an assessor has scored a candidate less than “average” for any competency or where wide discrepancies across a competency occur, an explanation is required of the assessment team. The assessment team also makes a recommendation as to the candidate’s suitability for the MP occupation. If the candidate is found unsuitable, the assessment team also indicates whether an unsuccessful candidate can re-apply for the MP occupation at a later date or another occupation within the CF. The Selection Board Chair makes the final recommendation as to whether the candidate should be accepted in the MP occupation.

As there is always a concern that observations can be influenced by assessor bias, attempts are made to minimize these potential effects, including:

a) Three-day training session for assessors prior to each MPAC with a goal of achieving inter-rater reliability;

b) Multiple inputs of scores from multiple assessors (11 competencies, 4-6 observers, 5 exercises);

c) Five-point rating scales with behavioural anchors for each of the eleven competencies;

d) Observation of candidates in a wide range of job simulation exercises, to ensure that each competency is tapped into more than once; and

e) Requirement of assessors to justify extreme scores to the Assessment Team Leader.

**MPAC Tools**

There are currently five tools included in the revised MPAC to evaluate the candidates. These selection measures are based on the newly revised competencies. The tools that form the assessment phase of the selection process follow as Table 2.

In order for a natural “hierarchy” to emerge and a true reflection of personalities to be shown, the group dynamics exercise is the first assessment exercise the candidates perform. This exercise is conducted the first evening the candidates arrive, after an introduction to the process given by the Selection Board Chair. The completion of the exercise so early in the process ensures that candidates have not become overly familiar with one another and instituted their own informal “pecking order”. The remainder of the exercises are completed in no specific order during the next day.

There are two additional measures included in the revised MPAC, but they are in developmental stages and are used for research purposes only. These include:
a) A personality measure (Trait Self-Descriptive Inventory), and
b) A diversity survey (portion of the CF Diversity Climate Survey).
Table 2

MPAC Exercises

<table>
<thead>
<tr>
<th>Exercise</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group Dynamics Exercise</strong></td>
<td>A leaderless group discussion is used in which the participants, working in small groups, work together to develop a consensus on how to handle a job related problem.</td>
</tr>
<tr>
<td>Skills Test</td>
<td>This test combines measures of a candidate’s observation skills and memory. In addition, it is designed to test the candidate’s ability to analyze a controversial topic, balance the argument, and provide convincing evidence in writing.</td>
</tr>
<tr>
<td>Background/Integrity Interview</td>
<td>A one-on-one semi-structured interview is used to probe into a candidate’s background and integrity.</td>
</tr>
<tr>
<td>Role Play Exercise</td>
<td>Candidates play the role of an MP in dealing with a citizen’s problem. The situations are typical of the more sensitive communications that a MP member must handle.</td>
</tr>
<tr>
<td>Structured Interview</td>
<td>A one-on-one interview that uses a series of structured situational and behavioural type questions to tap into specific competencies.</td>
</tr>
</tbody>
</table>

**CONCLUSION AND WAY AHEAD**

While the MP Branch saw the 2002 trial of the “new” MPAC as very successful, the research team still has plans to implement two more measures into the process. These two exercises include a “Fact-Find” exercise, which is essentially a crime scene role play that requires the candidate to write up a report of their observations, and a third role play on the issue of integrity. As integrity is the only competency considered by the MP Branch to be a “go/no go” factor, it was determined that it needed to be observed “in action” as well as through the Background/Integrity Interview. Once sufficient data is collected, a validation of the entire process will be conducted.
REFERENCES


For many years, ARI participated in IMTA meetings. Then, for reasons beyond the scope of this paper, we stopped. Now we are back, and it seems like an opportune time to take a historical look at our research on selection. I would like to look at where we were when our participation stopped, where we have gone since then, and where we think we are going in the future.

**PROJECT A/CAREER FORCE**

When we stopped attending IMTA, circa 1995, we were just finishing a 13-year effort which incorporated Project A and a related successor project, Building the Career Force. This effort was distinctive in terms of its scope as well as its length. It involved a multitude of predictor variables, including the Armed Services Vocational Aptitude Battery (ASVAB), a cognitive aptitude test battery used for operational selection and classification for all the military services, as well as psychomotor, spatial, interests, and temperament measures. It involved a wide variety of performance measures, including hands-on tests, job knowledge tests, ratings, administrative measures, and, as soldiers were assessed in their second tour of duty, situational judgment tests and simulations. It included a concurrent and a longitudinal validation. It advanced our understanding of performance, showing a major split between measures of “can do” and “will do” performance measures, and used this new understanding to provide a more complete picture of how individual difference measures related to performance. Our cognitive measure, the ASVAB, was a strong predictor of “can do” performance but was only moderately effective in predicting “will do” performance. Our temperament measures provided substantial incremental validity in predicting will do performance, particularly when the elapsed time between the administration of the temperament measures and administration of the performance measures was minimal (Campbell & Knapp, 2001).

**TEMPERAMENT MEASURES**

---

6 All statements expressed in this paper are those of the author and do not necessarily reflect the official opinions or policies of the Army Research Institute or the Department of the Army.
These findings stimulated considerable interest in the use of temperament measures in selection. A major hurdle to such use was the perceived problem of faking. Although opinions on the importance of faking varied, our findings indicated that when test takers were motivated to fake, they faked enough to potentially undermine the validity of our tests (Young, White, & Oppler, 1991). We pursued a number of approaches to counteract faking, including the use of biodata, a faking scale, and a forced-choice format. Finally, using the forced-choice format, we obtained findings sufficiently encouraging that we were able to conduct an operational test of our new measure (Young, Heggestad, Rumsey, & White, 2000). Early findings from this test indicated that further refinements were needed (M. C. Young, personal communication, January, 2002), and we have begun to make such refinements.

SPATIAL AND PSYCHOMOTOR MEASURES

Positive findings were also obtained for spatial and psychomotor measures administered in the Project A/Career Force projects. However, these measures provided limited incremental validity over and above the ASVAB for predicting can do performance (Oppler, McCloy, Peterson, Russell, & Campbell, 2001). In a number of separate research projects, it was determined that spatial and psychomotor tests did provide substantial incremental validity for predicting performance in certain gunnery tasks (Walker & Rumsey, 2001). Six of these tests were examined in the Department of Defense’s Enhanced Computer Administered Testing (ECAT) project, which examined the potential value of a variety of tests for enhancing the validity of the ASVAB. Assembling Objects was one of a number of measures which added to the validity of ASVAB (Wolfe, 1997), and because of this and a number of other appealing characteristics about it, a decision was made to include Assembling Objects in the battery administered to applicants in the joint service Military Enlisted Processing Stations, although the test has not been used for job assignment because the full scale validation research across all jobs needed to support such an application has not been conducted.

POST-ENLISTMENT SELECTION

Historically, our enlisted selection research has focused almost exclusively upon initial entry decisions. However, it is just as important for the Army to identify whom to tap for advanced assignments as whom to bring into the organization. The longitudinal design of Project A/Building the Career Force allowed us to begin exploring characteristics that would be important for advancement. Perhaps the most striking of the second tour findings was that the cognitive measures that predicted so well in first tour maintained their validity in predicting performance several years later (Rumsey, Peterson, Oppler, & Campbell, 1996).

One distinctive characteristic of Project A was the scope of individual differences examined. However, as theory on the complexity of cognitive ability advanced in the 1980s (e.g., Sternberg, 1985), we began to consider whether new measures might contribute to our ability to predict performance, particularly at the level of noncommissioned officer, or NCO. In Expanding the Concept of Quality of Enlisted
Personnel, or ECQUIP, we added a measure of tacit knowledge to measures of temperament and general cognitive aptitude in predicting ratings of NCO performance. The most powerful predictor was a composite temperament measure, the Assessment of Background and Life Experiences (ABLE), which was also such a powerful predictor of will do performance in Project A. The tacit knowledge measure modestly incremented general cognitive aptitude and ABLE as a predictor of rated performance (Peterson, et al., 1999).

**SELECTION FOR THE FUTURE**

Until recently, our orientation has been on selecting soldiers needed for current operations. However, as the Army has begun to confront a future of diverse operations, increasing uncertainty, multiple threats, advancing technology and other changes, it has raised the question of whether current selection procedures are applicable for future operations.

The concept of selection never exists in a vacuum. We are always selecting people to perform some job or set of jobs. If the organization we are placing individuals in is not changing rapidly, the job for which we are making the selection is likely to remain relatively stable. If the organization is undergoing rapid change, then the very meaning of what constitutes a job becomes more difficult to specify. A job today may not be the same as a job tomorrow. Tomorrow’s job may represent a consolidation of two or three of today’s jobs. Or it may be an entirely new job, designed to maximize the potential of some new piece of equipment or to achieve an entirely new type of mission.

Accordingly, we must give careful thought to how we analyze future jobs. Any analysis based on present job requirements is only relevant to the extent that these requirements are likely to remain relevant in the future. Inevitably, some projection will be needed based on judgment, ideally expert judgment. Projections which involve broad outlines of job requirements are likely to be safer than projections which involve specific tasks. Our approach, based on work done by Schneider and Konz (1989), has involved basing projections on available information about current job requirements and on expected changes in the environment that will impact upon these requirements. However, we have tended to focus on general job dimensions more than on specific tasks.

A shift in focus from present to future requirements does not relieve us of the responsibility to validate our proposed selection measures against job performance. However, it does make the problem of developing appropriate performance measures more difficult. We have identified performance dimensions based on expected future requirements. How do we get meaningful evaluations of these dimensions from those who can only observe present performance? One approach we have been pursuing is the use of future-oriented scenarios, based on our future-oriented job analyses (Hess, Entin, & Miller, 2002). If we can create reasonable representations of future environments and have raters observe performance in these environments, their ratings may contain valuable predictive information.
We sometimes encounter skepticism about whether a focus on the future will really tell us anything new about what we should be measuring for future selection. The skeptical view is that we already know what attributes contribute to successful performance and the projected changes in job requirements are not so dramatic as to make any difference in how we should select soldiers.

Is this a fair assessment? Let us see what the research we have conducted thus far suggests, particularly a project known as 21st Century NCOs, which is examining new ways of improving promotion measures for future NCOs. While the results of this project do not apply directly to initial entry selection decisions, since they deal with performance of soldiers at reasonably advanced levels, future projections do suggest that higher level responsibilities are being pushed down to lower levels, so these results are at least indirectly relevant to such decisions. The job analyses from this project suggest that the future environment will be very uncertain, and that units will need to be prepared to deal with a variety of potential threats. Accordingly, considerable emphasis was placed on judgment and decision making, particularly at higher levels and for the period extending beyond 2010 (Ford, R. C. Campbell, J. P. Campbell, Knapp, & Walker, 2000). Current selection is based on general cognitive aptitude. Is a test of general cognitive aptitude a good test of judgment and decision making? If the test includes mathematics and verbal and information items, as does the ASVAB, the answer is probably no, and the addition of a judgment test could provide new, useful information. A Situational Judgment Test was examined in NCO21, and was shown to be a better predictor of performance ratings than the ASVAB composite examined, General Technical (GT) (Sager, Putka, & Knapp, 2002).

Another attribute which is often linked to future job requirements is adaptability. Our NCO21 job analyses were somewhat inconclusive on the importance of this attribute, with psychologists citing its importance to a greater extent than military subject matter experts, and senior NCO performance expected to be more dependent upon this attribute than junior NCO performance (Ford, et al., 2000). One difficulty in exploring this attribute is its complexity. It can be viewed as a predictor linked to a “combination of cognitive, temperament and motivational factors (Rumsey, 1995)” as well as a multi-dimensional performance construct. Pulakos, Arad, Plamondon, and Kiechel (1997) identified eight different adaptability performance dimensions. On the predictor side, tolerance for ambiguity may be viewed as a cognitive style which facilitates adaptability. A measure of tolerance for ambiguity was found to be significantly related to rated performance in the NCO21 project (Sager, Putka, & Knapp, 2002).

Other attributes identified which appear to be particularly relevant to future environments were these: General Self-Management Skill, Knowledge of System Inter-Relations, Management of Battlefield Functions, and Self-Directed Learning (Ford, et al., 2000). Knowledge of System Inter-Relations and Management of Battlefield Functions seem to be attributes which would be particularly important at higher organizational levels; General Self-Management Skill and Self-Directed Learning are attributes which might be useful at any stage in a soldier’s career. Aside from Knowledge of System Inter-Relations, which was measured in a special form of the Situational Judgment Test,
none of these was directly measured as a predictor in the NCO21 project. However, positive findings for military education and civilian education as predictors might be viewed as supportive of the importance of Self-Directed Learning (Sager, et al., 2002).
CLASSIFICATION AND PERFORMANCE MEASUREMENT

In the Army, two decisions about an individual’s job status are made prior to entry. The first decision is whether or not the individual meets the minimal qualification for entry into any Army job. This decision is based primarily upon the individual’s score on an ASVAB composite of two verbal and two mathematics tests, the Armed Forces Qualification Test, or AFQT. If the individual achieves a sufficiently high score on this measure to qualify for Army entry, then the second decision must be confronted: which job or jobs is the individual most qualified for? This is the classification, or person-job match, decision.

Classification measures are used to place the individual where he or she is likely to perform in a way that will most benefit the organization. Tests which are highly predictive of performance in one set of jobs, but not predictive of performance in other jobs, are particularly useful for classification. To know whether classification tests are achieving their purpose, it is important to know how they relate to performance on each job in a particular organization. In an organization with as many different jobs as the Army, the development of performance measures that would be needed to cover every job would be extraordinarily expensive. For a period of time, the Army developed and maintained job-specific performance tests, known as Skill Qualification Tests, or SQTs, for training and assessment purposes, but this testing program has since been dropped.

Recently, Zeidner and Johnson have developed a procedure which has been used to improve the Army’s classification system. This procedure involved generating predicted performance scores from available validation data, then using a series of simulations to determine which multiple combinations of tests and jobs produce the greatest predicted performance across an entire cohort of new Army recruits (Zeidner & Johnson, 1994). It was possible because of the availability of historical SQT scores on individuals in most Army jobs.

In order to continue classification research into the future, we will need to either help the Army build cost-effective job-specific performance measures for its use and ours, or find a way to develop such measures to use in our research. It is possible that computerized testing may reduce administrative costs, but developmental costs will still be a major consideration. Possible strategies for reducing developmental costs include sampling from task or job clusters, or identifying components of work at more global levels than we have done traditionally. Both tasks and jobs are expected to undergo major change in the future, suggesting the need for an organizing scheme which is as flexible as possible.

CONCLUSION

The completion of Project A and Building the Career Force represented a landmark in the history of Army selection and classification research. However, it did
not mark the end of such research. As Army jobs and missions undergo revolutionary change, the need to examine new selection and performance measures keyed to this change has emerged as a critically important research issue.

REFERENCES


Revitalization of Navy Enlisted Selection and Classification

David L. Alderton, Ph.D., & Murray W. Rowe
Navy Personnel Research, Studies, and Technology Department (PERS-13)
Navy Personnel Command
5720 Integrity Drive, Millington, TN, USA 38055-1300
David.Alderton@persnet.navy.mil

US military enlisted selection and classification (S&C) processes were developed for conscripted service using early-20th-century manufacturing models. That is, large volumes of “raw material” were herded through test halls and group medical exams. The “raw material” was crudely sorted by quality and assigned to training regimes based modestly on quality but largely on near-term manpower requirements. The process presupposes nearly limitless input resources, is optimized for volume, and cares little about wastage. This was true in 1950 when conscription was reintroduced after WWII, and unfortunately, it is true today, more than 25 years after the enlisted ranks became an all-volunteer professional force.

Let me describe the current process. We bring candidates in for group mental testing with the Armed Services Vocational Aptitude Battery (ASVAB). This is a collection of 8 tests that measure verbal, math, and technical knowledge. We collect criminal, financial, and educational affidavits, followed by a medical examination. If the person passes the basic screens, they are sent to a classifier who is the first person who can discuss particular jobs. The classifier types in their Social Security Number, ASVAB scores, and medical, moral, and educational information and submits this to a text-based system designed in 1978 (CLASP). CLASP screens the candidate for minimal qualifications for the Navy’s top 15 manpower critical ratings. The classifier then proceeds to sell the candidate one of these jobs; the average interview lasts 7-10 minutes which determines their entire military career, including duty stations, promotion rates, special pays, and many of their civilian opportunities.

Although there can be little doubt of the value of aptitude testing for selection and classification (Hunter, 1983; Hunter & Hunter, 1984; Ones & Viswesvaran, 2002), there are many problems with this process. It does not coincide with the expectations of today’s youth, it is uninformative and asymmetric in terms of information; it is driven by manpower needs and discourages input from candidates. The candidate has little opportunity for input and even less realistic choice among decisions. Additionally, it does not foster a good match between the candidates’ abilities, interests, and desires and the full-spectrum of available Navy jobs. As a result, frequently a very bright candidate is placed in a job that is well below their intellectual ability and, late in the recruiting year, we often have to make too many exceptions (or waivers) to place less qualified people in unfilled but difficult jobs.

Set aside for the moment my concerns with how selection and classification is conducted. What indications are there that a problem exists? The turnover literature tells

7 The opinions expressed are those of the authors. They are not official and may not represent the views of the U.S. Navy.
us that if S&C systems are failing, we should expect much early attrition, academic training failures, high turnover, and poor retention rates (Hom & Griffith, 1995; Hom & Kinicki, 2001; Ryan, Sacco, McFarland, & Kriska, 2000). Despite over three years of concerted efforts in the US Navy, the most recent estimate of first-term cohort attrition is 37% (Center for Naval Analyses, 17 October 2002); i.e., almost 4 of 10 recruits will not complete their enlistment contract. In fact, between basic training and initial advanced training, more than 10,000 of 52,000 recruits will attrite at a cost of nearly $200,000,000 – before a single day of productive work for the Navy. We will lose another 9,000 from the Fleet, before the end of their obligated service contract. I argue that there are a number of problems with the current S&C system that leads us to this hemorrhage of personnel, and this symposium consists of projects that are attempting to mitigate the problems or increase our understanding of them.

One problem is that the current process does not seem fair, relevant, or to use the in vogue term, transparent, to applicants (c.f., Gilliland, 1993; Ryan, Sacco, McFarland, & Kriska, 2000). Many candidates feel that information is withheld and that they are not afforded any choice in the process. To address this, we developed the Rating Identification Engine (RIDE; which Dr. Farmer will discuss). Much of the work on RIDE was focused on developing a new assignment algorithm that provides a substantially better match between the individual and Navy jobs; RIDE is also better at overall resource allocation. Moreover, the underlying architecture of the algorithm is quite flexible and can accommodate new tests, qualifications standards, and changes in the Navy job structure. However, another important component of RIDE focused on the user interface design, with the explicit intent of improving the amount of information available about jobs for applicants. RIDE provides rank ordered lists of qualified jobs, including enlistment incentives, ship dates, and the like. Because of the amount of information presented, classifiers invariably turn the monitor so that the applicant can view the information, the two then discuss aspects of it, and the applicant may trade-off incentives for ship dates, or specific jobs, etc. The centrally important result is that RIDE has turned classification into something much more like a vocational counseling session where the candidate is participating. Early results from a live pilot test in San Diego indicate that RIDE classified candidates have lower Delayed Entry Program (DEP) attrition, lower basic training attrition, and lower A-school attrition (N=1158 RIDE vs. N=2312 CLASP classified candidates). Significantly, RIDE has been adopted by the Navy and will become operational in March 2003.
While the RIDE interface provides the opportunity for vocational discussion and for the applicant to make explicit trade-offs among jobs, based on incentives and ship dates, there is no formal mechanism for an applicant to express preferences among the jobs themselves. Why should the Navy care? The evidence is that classification decisions that accommodate preferences increase training success, job satisfaction, job performance, and retention (e.g., Barak, 2001; Ryan et al., 2000; Judge, Thoresen, Bono, & Patton, 2001). With entry-level Navy jobs, expressing preferences is complicated because naïve candidates do not understand the arcane names (e.g., Fire Control Technician) or technical terms (e.g., side-band radar) used to describe them. However, we have developed an instrument, Job and Occupational Interest in the Navy (JOIN), that allows naïve individuals to state their preferences using pictures and short (2-word) textual expressions; it takes about 20 minutes and produces a complete rank ordering of all 80+ entry-level jobs. We have conducted a preliminary study with JOIN and the results were very encouraging; Dr. Farmer will describe this further.

RIDE and JOIN ameliorate many problems with the current S&C system, yet there still remains the fundamental problem that we are only predicting training outcomes. That is, all of the statistical machinery is aimed at controlling training attrition, criteria which has a horizon of the first 12 weeks of a 208-week enlistment contract. The predictors that we use, ASVAB and high school diploma status, perform admirably for this, yet they cannot capture the complexities of military work life, particularly aspects of adjustment, team work, social judgment, and work styles which make or break an individual following training. As a result, we are undertaking several projects that will assess and develop non-cognitive measures, such as personality, social judgment, psychosocial adaptability, positive and negative affect, to determine which are most likely to improve both selection and classification, but with a focus on post-training outcomes such as adaptability, commitment, career success, and retention (Ackerman & Heggestad, 1999; Barrick, Mount, & Judge, 2002). Dr. Farmer will discuss one of these efforts.

There is a tacit assumption that attrition is solely a function of the individual and usually assumed to be a selection failure. Dr. Mottern will discuss a project called “1st Watch.” The philosophical orientation is that attrition is not the fault of the selection system or the individual, rather, it is a function of a mismatch between the person and the organization, and that the purpose of training is not just imparting knowledge, but rather to improve the fit between the person and the organization. First Watch consists of a series of surveys at important junctions throughout basic and initial advanced training, with the focus being to assess how well the individual fits within the organization (Person-Organization or P-O fit), how they adapt and change through the course of training and indoctrination, and how training fosters the development of core values and commitment. Attrition in this framework is considered a function of a mismatch between the expectations and characteristics of the individual and those of the Navy. This rich set of interconnected surveys provide insight into the “Sailorization” process, what our training is doing well and wrong, and how a better fit can be achieved. It also provides an alternative set of measures to evaluate the impact of changes in S&C and to predict subsequent attrition.
We must fundamentally reorient our enlisted S&C processes to become more applicant-centric, to better understand and value the applicant as an individual, just as a company would value a new employee. We must increase the breadth of information we collect about applicants to include not just achievement and aptitude, but also social and personal predispositions, as well as interests and preferences. We must utilize all of this information to make the best match between the individual and entry-level Navy enlisted jobs. And this must be done as a transparent process, one that appears open and fair to the applicant, and one in which the applicant fully participates in the decision, trading-off job attributes and incentives as they wish. Moreover, we must move our predictive horizon beyond initial training; to do this, we must begin to measure and predict outcomes such as job satisfaction, job success, adaptability, promotion rates, career potential, and retention. These outcomes are in many respects more relevant to both the individual and the Navy. By making selection and classification decisions from a broader array of information from the applicant, and by incorporating both near- and late-term outcomes, we move closer toward the ideal of a Whole Person Assessment; we improve our area of study, and most importantly, we improve the lives of Sailors and productivity and readiness of the Navy.

REFERENCES


Center for Naval Analyses (17 October 2002). *Attrition and Reenlistment of First-Term Sailors: Update Through 2nd Quarter FY02*. Briefing prepared for the Chief of Naval Operations.


1st WATCH ON THE FIRST TERM OF ENLISTMENT

Jacqueline A. Mottern, Ph.D.
Michael A. White, Ph.D.
David L. Alderton, Ph.D. 8

U. S. Navy Personnel Research, Studies and Technology, PERS-13
Millington, TN 38055-1300
P13h@persnet.navy.mil

This project, which we call 1st Watch on the First Term of Enlistment, represents an effort to better understand Sailors’ career progress during the important first term of enlistment and to use that information to help produce high quality sailors who are well prepared for their careers in the Navy and for life. This paper outlines the 1st Watch project, which began data collection in April 2002.

Background

This project is based on the idea of modifying traditional measures of person-organization fit for use with Navy recruits and Sailors in their first term of enlistment. Here, person-organization fit was defined as the similarity between individuals’ self-perceptions of personal characteristics and the Navy’s (organization’s) desired personal characteristics.

Person-Organization Fit Theory

Person-organization (P-O) fit has been shown to be related to attraction to an organization, organizational commitment, longer tenure, intention to quit, and actual turnover (O’Reilly, Chatman, & Caldwell, 1991; Bretz & Judge, 1994; Chatman, 1991; McCulloch & Turban, 2001). Most of the research uses the definition of P-O fit proposed by Kristof (1996): “The compatibility between people and organizations that occurs when (a) at least one entity provides what the other needs, or (b) they share similar fundamental characteristics, or (c) both.” Schneider, Goldstein, & Smith (1995) report support for the view that individuals are attracted to organizations that they believe provide a good fit with their personality, attitudes, and values. From an organizational perspective, Bowen, Leford, & Nathan (1991) suggest that organizations select individuals that they believe fit the values of the organization as well as the requirements of the job. Many researchers have focused on the match between people’s values and the values of the organization (Chatman, 1991). Organizational values were usually determined by incumbents’ ratings while individuals rated their preferences for working in organizations with this profile. The correlation between the two profiles is the usual operational definition of fit. The Organizational Culture Profile (OCP) (cf. Chatman, 1991; Karren & Graves, 1994; Kristof, 1996; O’Reilly et al., 1991; McCulloch & Turban,

8 The opinions expressed are those of the authors. They are not official and may not represent the views of the U.S. Navy.
(2001) is one such measure that has been used to assay the effects of person-organization fit on career decisions of employees in civilian occupations.

Instead of comparing the individuals’ preferences with the organization’s profile, as typically found in the literature, we derived the Navy’s desired profile for individuals from the Evaluation Report & Counseling Record (E1-E6) and compared that to individuals’ self-reports of the desired characteristics. We call this new measure of person-organization fit the Navy Fit Scale. We intend to use this new measure, along with others, to predict retention and attrition during the first term of enlistment.

More specifically, we extend existing research on person-organization fit by:
(1) developing an organizational profile of desired Sailor characteristics and comparing individuals’ self-perceptions with that profile; (2) examining the effectiveness of P-O fit in predicting retention and attrition during training; and (3) investigating changes in P-O fit as a function of stage of training. Ultimately, our purpose is to improve the predictors of long-term retention and attrition in the Navy.

Measures

For this project, we have developed a new measure of person-organization fit and condensed existing measures of organizational commitment and stress coping skills to meet project needs. Additional measures of recruit expectations, social support by family and peers and training staff during training, and training experiences are also used.

**Navy Fit Scale**

We developed a set of statements regarding the Navy’s desired characteristics of first term Sailors from the Navy’s Evaluation Report & Counseling Record (E1-E6). This form has seven “blocks” or groups of desired behavioral characteristics and Sailors receive annual ratings on all the blocks. Scores range from “1” (below standard) to “5” (exceeds standard) with “3” as average. We chose six of these blocks, excluding only the block measuring job expertise, and wrote statements reflecting the average. Three groups of senior enlisted Sailors (N=18) used a Q-sort process to identify the statements that represented the most desired characteristics of new Sailors for each block. Items were then pre-tested on a sample of new recruits (N=186) and factor analyzed. Six factors emerged representing the theoretical blocks of the Evaluation Report and Counseling Record (E1-E6) and have coefficient alphas ranging from .67 to .82. The Navy Fit Scale contains the following six subscales: Quality of Work, Respect for Others, Military Bearing/Character, Job Initiative, Teamwork, and Leadership. Sailors are asked to use a 5-point Likert type scale to indicate the degree to which they have a list of personal characteristics. The difference between the Navy’s desired individual characteristics and the individuals’ profile is the operational measure of fit.

**Navy Commitment Scale**

Organizational commitment has been conceptualized and measured in various ways, but common to all is the link between commitment and outcome; employees who are strongly committed are least likely to leave an organization. We used a model of commitment developed by Meyer & Allen (1987) that includes three components: affective, continuance, and normative. Using the 24-item scale (with slight re-wording to
reflect Navy terminology) plus additional items we developed to measure Navy value similarity, we pre-tested the scale on more than 800 Navy personnel. After factor analyzing the results, we retained four items from the Affective Commitment Scale (ACS), three items from the Continuance Commitment Scale (CCS), and five items from the Value Similarity Scale (VSS). The Normative Commitment Scale (NCS) was dropped due to range restrictions and low factor loadings across 3 additional factors. The ACS, CCS, and VSS had alphas of .78 or higher and were related to reported career intent in the pre-test. We have a paper in progress on the psychometric qualities of this measure, which we call the Navy Commitment Scale.

**Sailor Stress Coping Skills Scale**

Beginning a career in the Navy can be a very stressful event in the lives of our young recruits. Leaving home, family and friends to live with 60-80 strangers is in itself stressful for most recruits. Living in a barracks for eight weeks with these strangers while having all your time closely supervised and evaluated by a Recruit Division Commander (RDC) can only increase the stress. As noted by Cable and Parsons (2001) “Stress and anxiety are likely to be high when newcomers find themselves in a new organization because they do not possess comfortable routines for handling interactions and predicting the responses of others.” We included the revised Ways of Coping Checklist (WCCL) developed by Vitaliano, Russo, Carr, Maiuro & Becker (1985) in our pre-test. The 42-item scale factored into the same five factors as did Vitaliano, et al., but we eliminated 10 items that did not perform well in our pre-test (factor loadings below .30 and scattered across the primary factors). We believe that recruits with better stress coping skills will be more likely to complete training than those with weaker skills. We hope to identify the skills most closely related to training success and use this profile for future recruit training.

**Retention**

On average, the Navy has approximately 65% of new recruits complete their first term of service. A smaller percent will re-enlist (sign a new contract extending their service) and serve beyond their first term. Based on prior research, we believe that better P-O fit, higher commitment, and stronger stress coping skills will all be positively related to retention.

**Survivability**

Attrition in the Navy is traditionally defined as failure to complete the first term of obligated service. Rather than look at attrition as a dichotomous variable (completing or not completing the first term), we propose to examine attrition rate as a continuous variable measured by weeks of service. In practical terms, the more quickly the Navy losess a Sailor, the smaller our return on investment. After investing considerable time, effort, and money in recruiting and training an individual, it is in the Navy’s best interest to have a productive Sailor as the outcome. We define how quickly a recruit leaves the Navy as “survivability” and theorize that low levels of P-O fit, low commitment, poor stress coping skills, low levels of perceived social support, unmet expectations of the Navy, and negative experiences in training are related to survivability.
Method

The Navy maintains recruiting offices across the country and active duty Navy personnel are assigned to recruiting duties. Each recruiter is given a monthly goal of recruits to enlist and must work diligently to meet that goal. Applicants are screened and tested to ensure they meet Navy qualification standards. Once this process is complete, each applicant meets with a Navy classifier who knows the Navy’s current manpower needs and specific job availability. That information, combined with applicant test scores and job interests, result in a contract specifying a rate (job series) and, possibly, training schools that are promised the applicant. Applicants are given a date for reporting for service and travel to Great Lakes Navy Training Center for their initial Recruit Training Course.

Unless coming from the immediate area, recruits fly into O’Hare airport and report to a holding area. Each evening, Navy personnel from Great Lakes arrive to escort charter buses of new recruits to Great Lakes. At Great Lakes, recruits begin a tightly controlled in-processing that includes the traditional haircuts, new uniform items, and shots. From in-processing, recruits are assigned to a training division and begin eight weeks of training. If trainees successfully complete training requirements, they graduate from RTC and go to the next phase of training.

For those promised advanced training for a rate, an advanced or “A” school waits. About half of the Navy’s “A” schools are located at Great Lakes. The remainder travel to a different location for advanced training. After completing “A” schools, some sailors go on to even more advanced schools before reporting to their first assignment. Others report directly to the fleet. For those RTC grads not promised advanced schools, the Navy operates three apprentice schools at Great Lakes. These schools last 2-3 weeks and ship these sailors to their first assignment in the fleet.

During RTC, “A” schools and Apprentice schools, trainees attrite for a variety of reasons: medical problems (such as injury during training), drug use, misconduct, and problems that existed prior to enlistment.

Subjects

A one-year cohort of recruits who begin training from April 2002 through March 2003 will compose the sample (N approximately 50,000). This cohort will be followed through their first term of enlistment, normally four years.

Procedures

Data from this project will be collected at five points across four years, as described below. Four surveys were developed for this project and linked to the existing Argus survey for tracking purposes.

New Sailor Survey

The first questionnaire is being administered on the first day of in-processing. All new recruits are asked to complete the survey while in transit to Great Lakes Navy Training Center. Surveys are passed out by a Navy Petty Officer and collected upon
arrival at Great Lakes. The questionnaire assesses individuals’ personal values, their experiences with recruiting and classifying, their reasons for joining the Navy, their expectation of training, their stress coping skills, the Navy Fit Scale and demographic information. As of October 4, 2002, data from approximately 19,455 recruits have been processed.

**RTC Graduate Survey**

Of the sailors who completed the New Sailor Survey, 9,447 have completed the RTC Graduate Survey. This questionnaire is administered to all trainees who are on the 1st day, 8th week of RTC and identified for graduation. Questionnaires are administered by Petty Officers in a classroom prior to an informational briefing on benefits. Completed questionnaires are dropped in a box, which is then sealed and shipped to our data processing center. The Navy Fit Scale, Navy Commitment Scale, training experiences, and an evaluation of RTC are included.

**“A”/Apprentice Graduate Survey**

Sailors are surveyed in the last week of training for their advanced or apprentice school. The student class leader distributes the questionnaire, collects them and returns them to a school official. Surveys are boxed and returned to our data processing center. This questionnaire includes all of the items in the RTC graduate survey, but is updated to apply to specific events in advanced or apprentice training. To date, 602 “A”/Apprentice school graduates have completed the questionnaire.

**Exit From Training Survey**

All sailors who leave the Navy during training are asked to complete a questionnaire at the separation barracks while awaiting discharge. To date, 735 have completed a questionnaire. Prior to a legal brief, Navy personnel assigned to the separation barracks distribute the questionnaires, then box and return them to our data processing center. The Navy Fit Scale, experiences in the Navy, reasons for leaving and training experiences are included.

**Argus Career Milestone Tracking System**

Argus, the Navy’s web-based transition survey, will be used to track individuals’ career progress through their first term. Argus is completed at transition points, such as changes in location, promotions, and leaving the Navy.

**Discussion**

The Navy’s 1st Watch project promises a wealth of data on a cohort of recruits as they pass through their first term of enlistment. The longitudinal research design coupled with multiple data points and new measures will provide the Navy with in-depth information on our new Sailors. We hope to develop a profile of successful new Sailors at various stages of their careers and use this information to improve our accessions process.
References


ENGINEERING AND COMPUTING WORKFORCE TRENDS IN DND AND CF

Karen E. Daley & Megan E. Dewar
Director Strategic Human Resources
National Defence Headquarters, Ottawa, Canada K1A 0K2
Daley.KE@forces.gc.ca

ABSTRACT

This report presents the results of a study of Engineers and Computer Scientists (E&C) in the Department of National Defence (DND) and the Canadian Forces (CF) in 1990 and 2000. Demographic analyses show that civilian and military E&C workforces have decreased in size in the past decade, as have DND/CF populations overall. The proportion of women E&C DND/CF, and in the private sector, is smaller than the proportion of women in the Canadian workforce at large. The mean age and years of service of civilian and military E&C have increased since 1990. Separation rates have remained quite stable since 1990 but are higher than private sector rates. The ageing population and private sector competition underscore the importance of recruitment and retention for DND/CF E&C. Consequently, many recruitment and retention programs and incentives are offered to civilian and military E&C.

INTRODUCTION

The changing work environment – particularly the rapid and widespread application of innovative technologies to military and industrial operations – and changing Canadian demographics will likely have an impact on the operations of the Canadian Forces (CF) and the Department of National Defence (DND) in the next twenty years. Consequently, the role of Engineers and Computer Scientists (E&C) in a highly skilled, technologically advanced force and support structure will become increasingly important to position DND/CF to meet strategic objectives. In the private sector, the demand for these kinds of skills will likely increase in a similar way. This increasing demand will occur in the context of an ageing workforce and declining share of youth in the labour market (Lowe, 2001). Competition for scarce skills is arising and underscores the importance of recruiting and retaining service members and civilians in key areas such as E&C. Thus, the aim of this report is to examine the demographic profile of E&C in the DND/CF in 1990 and 2000 and identify challenges for recruitment and retention.

Methodology

DND civilian and military E&C groups for 1990 and 2000 were identified based on job descriptions that included engineering, computer systems, and engineering support. Thus, the following civilian occupational groups were included: Computer Systems Administration (CS), Engineering and Scientific Support (EG), Electronics (EL), and Engineering (EN-ENG). Only civilians in the long-term or permanent workforce (i.e., indeterminate, seasonal or term-greater-than-three-months) were included in analyses. Officers in the following military occupation codes (MOCs) were included⁹:

⁹ In 1997/98, MOC 45 was split into 24 and 46.
Engineer (24), Aerospace Engineer (41), Communications and Electrical Engineer (42), Land Electrical and Mechanical Engineer (43), Maritime Engineer (44), Military Engineer (45), and Airfield Engineer (46). Because of the differences in the makeup of the civilian and military E&C groups (e.g., broader scope of occupations in the civilian group, fundamental differences between civilian and military careers, etc.), direct comparisons between the two are not permissible. Selected information is provided on private sector E&C for context.

Data for DND/CF were obtained from a variety of sources, including databases maintained by DND/CF Human Resources research units. Because of the way data were captured, civilian data reflect end of calendar year (31 December) and military data reflect end of fiscal year (31 March). Information on private sector E&C was obtained from government sources, and published studies.

Descriptive demographic data on population figures, employment equity, age, retirement eligibility, years of service, and separation/attrition rates were calculated. Systematic analyses of historical trends were not conducted for this project. It should be noted that some records are incomplete due to missing information. As a result, when dividing the population based on a given criterion, the sum of the parts may not be equal to the total.

RESULTS

Demographic Profile

Between 1994 and 1998, the Federal Government reduced the size of the Federal Public Service (FPS) through a series of programs designed to reduce public spending. Consequently, the number of civilians at DND decreased by 48% (33,481 employees in 1990 vs. 17,305 in 2000). Over the period 1992 to 1997, the CF undertook the Forces Reduction Program (FRP). The voluntary retirement program had a great impact on the size and demographics of the Forces. The number of officers in the CF decreased by 33% (18,991 officers in 1990 vs. 12,801 in 2000). By comparing data from 1990 and 2000 the years of downsizing are taken into account, we look at a time significantly before and a time significantly after the reduction period.

Civilian E&C

The size of the engineering and computing workforce was 2508 in 1990 compared to 2303 in 2000. Table 1 shows the number of employees by occupational group. The 8% decrease for E&C occupations, although substantial, was not nearly as dramatic as the 48% decrease for DND overall from 1990 to 2000. This decrease would have been greater for the E&C group without the large increase in the number of Computer Systems Administration (CS) employees, which occurred throughout the federal government (Treasury Board Secretariat, 2002).
Table 1: Number of Employees by Occupational Group and Year

<table>
<thead>
<tr>
<th>Occupational Group</th>
<th>Number of Employees</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1990</td>
</tr>
<tr>
<td>Computer Systems Administration (CS)</td>
<td>603</td>
</tr>
<tr>
<td>Engineering and Scientific Support (EG)</td>
<td>980</td>
</tr>
<tr>
<td>Electronics (EL)</td>
<td>472</td>
</tr>
<tr>
<td>Engineering (EN-ENG)</td>
<td>453</td>
</tr>
<tr>
<td>TOTAL</td>
<td>2508</td>
</tr>
</tbody>
</table>

Military E&C

In 1990 the number of E&C in the CF was 4904. In 2000 this figure had dropped to 3138. While this is a significant reduction in strength, it is a trend evidenced across the entire CF. E&C represented 26% of CF officers in 1990. Table 2 shows the population of CF E&C by year and MOC. In 2000 this figure was essentially unchanged at 25% of the total officer population, hence similar reduction percentages in E&C and the officer corps. The CF E&C reductions of 35% were comparable to 33% for officer corps in the same time frame.

Table 2: Number of CF E&C Members by MOC and Year

<table>
<thead>
<tr>
<th>MOC</th>
<th>1990</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engineer (24)</td>
<td>N/A</td>
<td>428</td>
</tr>
<tr>
<td>Aerospace Engineer (41)</td>
<td>1235</td>
<td>684</td>
</tr>
<tr>
<td>Communications and Electrical Engineer (42)</td>
<td>1286</td>
<td>816</td>
</tr>
<tr>
<td>Land Electrical and Mechanical Engineer (43)</td>
<td>593</td>
<td>339</td>
</tr>
<tr>
<td>Maritime Engineer (44)</td>
<td>1037</td>
<td>685</td>
</tr>
<tr>
<td>Military Engineer (45)</td>
<td>753</td>
<td>2</td>
</tr>
<tr>
<td>Airfield Engineer (46)</td>
<td>N/A</td>
<td>184</td>
</tr>
<tr>
<td>TOTAL</td>
<td>4904</td>
<td>3138</td>
</tr>
</tbody>
</table>

Employment Equity Profile

Civilian E&C

In 1990, 10% of E&C were women; this figure increased to 17% by 2000. With respect to other employment equity designated groups, data are not available for 1990. However, when employees self-identified in 2000, 0.6% of the E&C community in DND identified themselves as being an Aboriginal person, 4.0% a member of a visible minority, and/or 2.4% a person with a disability.

Table 3 shows the actual and expected number of women, Aboriginal people, members of visible minorities, and persons with disabilities by occupational group for the E&C population. The expected number of these designated group members is calculated based on their representation (by occupation and geographical location) in the Canadian workforce at large, based on information provided by the federal government. In general, women were under-represented in the EG, EN-ENG, and EL groups relative to workforce availability. Members of visible minorities were also substantially under-represented in the EG and EL groups. It should be noted that members of visible minorities are under-represented in the FPS overall (Taskforce on the Participation of Visible Minorities in the Public Service, 2000). Persons with a disability were slightly under-represented in the
EG and EN-ENG groups. Aboriginal peoples were not under-represented in any E&C groups.

Table 3: DND Representation of Employment Equity Designated Groups in the E&C Community (based on 30 September 2001 population)

<table>
<thead>
<tr>
<th>Group</th>
<th>CS</th>
<th>EG</th>
<th>EN-ENG</th>
<th>EL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total in Group</td>
<td>1029</td>
<td>745</td>
<td>352</td>
<td>290</td>
</tr>
<tr>
<td>Actual</td>
<td>Expected</td>
<td>Actual</td>
<td>Expected</td>
<td>Actual</td>
</tr>
<tr>
<td>Women</td>
<td>294</td>
<td>283</td>
<td>97</td>
<td>348</td>
</tr>
<tr>
<td>Visible Minorities</td>
<td>72</td>
<td>85</td>
<td>13</td>
<td>39</td>
</tr>
<tr>
<td>Persons with a Disability</td>
<td>36</td>
<td>24</td>
<td>35</td>
<td>44</td>
</tr>
<tr>
<td>Aboriginal Peoples</td>
<td>13</td>
<td>5</td>
<td>14</td>
<td>7</td>
</tr>
</tbody>
</table>

Military E&C

Military employment equity data presented were obtained from the 1995 Diversity Survey (Chouinard & Chiasson, 1995). The gender mix of CF E&C is presented as the percentage of female members of this group (see Table 4). In 1990, the percent of female E&C was 6% and 10% in 2000. These figures are noticeably lower than the gender mix of the general CF officer population. In 1990 10% of officers were female, in 2000 13% of officers were female.

Table 4: CF Percentage of Women E&C by MOC and Year

<table>
<thead>
<tr>
<th>MOC</th>
<th>1990</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engineer (24)</td>
<td>N/A</td>
<td>7.2</td>
</tr>
<tr>
<td>Aerospace Engineer (41)</td>
<td>7.5</td>
<td>12.6</td>
</tr>
<tr>
<td>Communications and Electrical Engineer (42)</td>
<td>7.4</td>
<td>8.8</td>
</tr>
<tr>
<td>Land Electrical Engineer (43)</td>
<td>5.7</td>
<td>8.3</td>
</tr>
<tr>
<td>Maritime Engineer (44)</td>
<td>2.9</td>
<td>8.3</td>
</tr>
<tr>
<td>Military Engineer (45)</td>
<td>5.2</td>
<td>N/A</td>
</tr>
<tr>
<td>Airfield Engineer (46)</td>
<td>N/A</td>
<td>12.5</td>
</tr>
</tbody>
</table>

Table 5 shows the actual and expected number of Aboriginal people and members of visible minorities by MOC for the E&C population as of end March 1995. The expected figures are determined in the same way as those for the civilian population.

Table 5: Representation of Employment Equity Designated Groups in the E&C CF Community (based on end March 1995 population)

<table>
<thead>
<tr>
<th>MOC</th>
<th>24</th>
<th>41</th>
<th>42</th>
<th>43</th>
<th>44</th>
<th>45</th>
<th>46</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Population</td>
<td>N/A</td>
<td>1013</td>
<td>1219</td>
<td>472</td>
<td>883</td>
<td>733</td>
<td>N/A</td>
</tr>
<tr>
<td># Visible Minorities</td>
<td>N/A</td>
<td>20</td>
<td>28</td>
<td>12</td>
<td>27</td>
<td>17</td>
<td>N/A</td>
</tr>
<tr>
<td>Expected # Visible Minorities</td>
<td>N/A</td>
<td>91</td>
<td>110</td>
<td>42</td>
<td>79</td>
<td>66</td>
<td>N/A</td>
</tr>
<tr>
<td># Aboriginal Peoples</td>
<td>N/A</td>
<td>9</td>
<td>8</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>N/A</td>
</tr>
<tr>
<td>Expected # Aboriginal Peoples</td>
<td>N/A</td>
<td>30</td>
<td>37</td>
<td>14</td>
<td>26</td>
<td>22</td>
<td>N/A</td>
</tr>
</tbody>
</table>
The data in Table 5 suggest that Aboriginal and visible minorities were under-represented in all MOCs relative to workforce availability. This pattern reflects the general CF officer population. In 1995, Aboriginal peoples made up 0.71% of the CF officer population versus 3% of the Canadian population. Visible minorities made up 1.96% of the CF officer population versus 9% of the Canadian population.

**Canadian Workforce**

E&C has been a non-traditional area of work for women. In 1990-91, Canadian women accounted for less than 4% of the registered engineers and received only 5% of the bachelor’s degrees awarded in the physical sciences and engineering (Sherriff & Binkley, 1997). However, the number of women enrolled in engineering degree programs in Canada has been increasing (Canadian Coalition of Women in Engineering, Science and Technology, 1997). To this end, there are many national and provincial organizations, universities, and professional associations working towards increasing the participation of women in E&C fields.

**Age Profile**

**Civilian E&C**

The mean age of the civilian E&C population in 1990 was 43 years, rising slightly to 44 years in 2000. CSs tended to be six years younger, on average, than the rest of the E&C group (mean age of 38 in 1990 vs. 41 in 2000). In contrast, the mean age of EGs, ELs and EN-ENGs together was 44 in 1990 and 47 in 2000. The average age of overall civilian population increased slightly from 1990 to 2000 (43 years vs. 45 years, respectively).

The age profile of the E&C group is also revealing. Figure 1 shows that although the average age has not changed much in the past decade, the proportion of employees aged 40 to 54 has increased (41% in 1990 vs. 60% in 2000). In 1990, 27% of the E&C population was less than 35 years (see Figure 1). By 2000, this percentage decreased to 14%.

This concentration of employees aged 40 to 54 is reflective of the trend for the department as a whole (Assistant Deputy Minister [Human Resources - Civilian], 2001) and the FPS (Lowe, 2001). The departmental increase in the 40 to 54 cohort is largely due to the combined effects of a hiring boom in the 1970s and workforce reductions, hiring freezes in the 1990s, and the ageing of the Baby Boom generation (born between 1947 and 1966).

**Military E&C**

The mean age of E&C in the CF was 32 in 1990 and 36 in 2000. It should be noted that the CF mandatory retirement age is 55, with few exceptions (Canadian Forces General Messages, 2001). Across E&C MOCs the mean age in 1990 ranged from 33 for Maritime Engineer to 34 for Military Engineer. In 2000 the mean age ranged from 34 in Engineer to 39 in Airfield Engineer. This ageing trend is reflective of the situation in the entire CF officer population (mean age 33 in 1990 vs. 36 in 2000). The percentage of the E&C population aged less than 35 years was 65% in 1990 and 42% in 2000. As with the civilian E&C, the downsizing of the mid-1990s contributed to this shift in demographics.
Figure 1: Civilian E&C Population by Age Group in 1990 and 2000

Figure 2: Military E&C Population by Age Group in 1990 and 2000

**Canadian Workforce**

The Canadian workforce is ageing. Lowe (2001) reports that by 2010, the retiring cohort (aged 54 to 64) will outnumber the youth share of the workforce (aged 15 to 25 years). Furthermore, as retirements increase, competition for younger workers will also
increase. The youngest cohort in the workforce (aged 15 to 25 years) is decreasing in size: this cohort comprised approximately 20% of the working age population in 1998 compared to about 30% in the late 1970s (Lowe, 2001).

In a sample of 13 Canadian companies with involvement in science and/or technology and investment in R&D, the mean age in each company ranged from 28 to 42 years (Charette, 1997). The mean age was lowest in areas where a shortage of workers exists, such as advanced technology and some areas of engineering. The average age tended to increase in areas where demand was not as high.

**Average Years of Service**

**Civilian E&C**

The mean YOS in the E&C community was 11 in 1990 and 14 in 2000. These averages reflect years of continuous service. The CS group had the lowest average YOS in 2000 at 11 years compared to the EL group (15 years) and the EN-ENG and EG groups (16 years).

**Military E&C**

The mean YOS for CF E&C was 12 in 1990 and increased to 15 in 2000. Across members of the E&C MOCs the mean YOS in 1990 ranged from 10 for Maritime Engineer to 14 for Communications and Electrical Engineer. In 2000, the mean YOS ranged from 14 for Engineer to 16 for Communications and Electrical Engineer. This increased YOS trend is reflective of the situation in the entire CF officer population. In 1990 the average YOS of serving officers was 12 versus 15 in 2000, the same as the E&C officers in both 1990 and 2000.

**Percentage Eligible to Retire**

**Civilian E&C**

For the purposes of this study, retirement eligibility is defined as the point at which civilians are eligible to retire with an unreduced annuity (generally at 30 years of service and minimum age of 55 or at age 60 with minimum 2 years of service). The percentage of employees eligible to retire decreased slightly (8% in 1990 vs. 5% in 2000). The observed decrease may be misleading for two reasons. First, large numbers of employees retired early with incentives during workforce reductions in the 1990s (Director Civilian Employment Strategies, 1995). Retirements are likely to increase in coming years as the oldest members of the Baby Boom generation (born between 1947 and 1966) approach retirement age. The EL and EN-ENG occupational groups have more than 40% of employees predicted to retire by 2011 (ADM [HR-Civ], 2001). In DND, large numbers of civilians will be eligible for retirement in 2006, but the peak is predicted to start in 2011 (Daley, 2001).

**Military E&C**

It is difficult to determine a definition of retirement eligibility for CF members as mandatory retirement from the CF occurs at age 55. Very few members remain in the CF until this age. Once members achieve 35 YOS they may retire with full pension. However, it is difficult to accrue 35 YOS before age 55, thus the majority of members leave the CF without a full pension. Both of these calculation methods yield very small
numbers of members considered eligible to retire. The percent of CF E&C eligible to retire due to age was 1% in both 1990 and 2000. An alternative way to consider the percentage of E&C eligible to leave the CF is to determine the number of service members at the end of their Indeterminate Engagement or IE (generally 20 YOS) when they are eligible for an immediate annuity. In 1990, 2% of E&C had 20 YOS compared to 3% in 2000.

Separation/Attrition Rates

Civilian E&C
Separation and attrition rates were calculated as the number of indeterminate (permanent) E&C employees who left in a given year as a percentage of the population at the start of the year. The separation/attrition rate for 2000 was 6% compared to 4% for 1990. Between 1990 and 2000, departure incentives were offered across the FPS. As a result, many civilians retired early, thus decreasing the available pool of employees eligible to retire in 2000, and possibly deflating the separation rate for this timeframe.

Military E&C
Military separation/attrition rates were determined using a database of released members. The overall attrition rate of E&C was 6% in 1990 and 2000. The mean age of released E&C was 36 years in 1990 and 40 in 2000. The mean YOS of released E&C was 16 years in 1990 and 19 in 2000. Data on reasons for release such as retirement or end of contract are maintained by the CF. The reasons most often quoted on release by CF E&C members (i.e., by at least 10% of releases in 1990 or 2000) were “On request – Entitled to immediate annuity, 18% in 1990 vs. 29% in 2000), “On request – Other causes, 38% in 1990 and 18% in 2000), and “On reaching retirement age, 26% in 1990 and 22% in 2000). Of the latter reason, over 60% were aged less than 54 in 1990 and 2000 and were not considered in the calculation of those who are retirement eligible.

Canadian Workforce
The attrition rate in a sample of 13 E&C companies was approximately 3% per year (Charette, 1997). Private sector E&C appear to have attrition rates than DND/CF E&C. One possible reason for the difference between the private sector compared to civilian E&C may be the higher mean age of civilian E&C. The reasons for the higher attrition rates for DND/CF E&C are not clear and without a controlled study, the exact reasons for differences in attrition rates cannot be determined. Nevertheless, the pattern of results does suggest that retention may be a more important issue to the CF and DND than to the private sector.

CONCLUSIONS
Implications of Results for Recruitment and Retention
The increasing demand for E&C stemming from the rapid pace of technological change combined with an ageing workforce and labour market means that the DND/CF will face increasing competition for scarce skills. Consequently, recruitment and retention of E&C workers for DND/CF will grow in importance in the years to come.
**Civilian E&C**

Recruitment and retention bonuses are being offered to E&C employees in the CS and EN-ENG groups. Work-life balance is also an important retention factor. According to a 1999 survey, 90% of DND employees in R&D reported that they are allowed the flexibility to balance their personal, family and work needs (Public Service Employee Survey, 1999). A recent study of executives, scientists, and other professionals in the federal Public Service revealed factors related to retention of these employees (Duxbury, Dyke & Lam, 1999). For example, job satisfaction was a more important retention factor than compensation. Furthermore, 96% of scientists and professionals indicated that the most important achievements would be for them to do work they enjoy and to gain a personal sense of accomplishment. The most common reason given (45% of respondents) for potential departure of an employee was unrewarding work, followed by better compensation (31%). Together these results suggest that public servants stay because of interesting, challenging jobs; conversely a frustrating job and the higher pay of the private sector would provide a strong incentive to leave the FPS.

**Military E&C**

A wide variety of incentives are in place to recruit and retain members of the CF in light of “significant recruiting and retention challenges” arising from the steady growth of the Canadian economy, reduced unemployment, and increased competition for young, skilled Canadians (Chief of the Defence Staff, 2002). Several officer occupations are in demand, including Maritime Engineer, Engineer and Aerospace Engineer and the CF initiated a program in 2001 to attract skilled technical workers with a signing bonus.

**Canadian Workforce**

Recruitment practices vary within the private sector. Charette (1997) reported on the practices in a sample of Canadian E&C companies: campus recruitment campaigns, student work terms, industry or lab visits, career fairs, the Internet, referral bonuses, foreign recruitment campaigns, and advertising. Charette described retention factors identified by the E&C sample: cutting-edge R&D, flexible employee benefits, permanent employment, parallel career paths (i.e., management or technical expert), salary, company culture, training and development, awards and recognition, communications, job mobility, and flexible work arrangements; stock options are also used (Hynes & Lendvay-Zwickl, 2001).

**Comparison of Recruitment and Retention Strategies**

With respect to compensation, the private sector can offer highly paid jobs, signing bonuses, retention bonuses and other financial incentives, often at the discretion of the hiring manager. Compensation in the DND/CF is generally predetermined and inflexible, as it is determined through collective bargaining or statute. With respect to hiring new recruits, private sector jobs can be offered very quickly whereas public service jobs can take many months to staff (ADM [HR-Civ], 2001). With respect to offering rewarding jobs, both the public and private sector offer diverse jobs in cutting-edge R&D, engineering and computer systems development and application. Civilian and military E&C jobs are more secure than jobs in the private sector, which is subject to
economic fluctuations. The FPS and the CF offer a wide array of medical, dental, death, and pension benefits, which are often better than in the private sector.

**Directions for Future Research**

Additional research to identify critical recruitment incentives and retention tools would allow DND/CF to focus resources on key areas. Studying employees’ and service members’ sources of job satisfaction or dissatisfaction would provide insights into reasons and consequently potential preventive measures. Quantitative data from controlled studies could be used to determine systematic variability in retention factors by E&C subgroups; qualitative data, such as interviews, could flesh out quantitative results. With this information, policy makers could develop strategies to improve recruitment and retention.

**Concluding Remarks**

This demographic analysis of civilian and military workforces in 1990 and 2000 has shown that the E&C workforces are ageing, as is the Canadian population overall. This trend places increasing importance on succession planning, recruitment and retention, particularly in light of competition from the private sector for highly skilled E&C. DND, the CF, and the private sector are competing for highly skilled, technologically advanced workers. Recent research suggests that the keys to retaining scientists and professionals, including E&C, are to provide interesting, challenging work with competitive compensation and to minimize the impediments to the conduct of this work. Future research could isolate the critical factors relating to recruitment and retention for DND/CF E&C groups.

**REFERENCES**


The Cross Cultural Adaptability Scale: A Psychometric Evaluation

Lieutenant(N) M. A. Vanderpool
Director Human Resources Research and Evaluation
National Defence Headquarters, Ottawa, Canada K1A 0K2

In a study of the perceptions of Australian peacekeepers, Schmidtchen (1997a) highlighted the importance of selecting suitable personnel for peacekeeping operations. Peacekeepers, Schmidtchen maintained, must demonstrate restraint in the use of force, act impartially, and resolve conflicts through mediation (1997a). Such requirements, however essential, can be extremely difficult to achieve. It is not unreasonable to expect, for example, that peacekeepers and officials from non-governmental organizations (NGOs) of any given contingent will frequently draw on their own cultural antecedents for reference points about appropriate behaviours in such situations. As will those of other contingents and those who are party to the hostilities or conflicts.

Unless carefully managed, interventions by peacekeeping forces may do more harm than good, and not just to the opposing factions. Peacekeepers’ own cultural standards of appropriate behaviour, for example, will frequently differ from those of members of the opposing factions, and even from those of other forces participating in the mission. Issues arising out of such differences may cause irreparable damage to the peacekeeping mission should they, for example, result in a decision that offends the cultural or religious sensitivities of either or both opposing parties, or even of another contingent. Failures or deficiencies in this regard can have a disastrous impact at a mission, force, national, or even international level.

A tragic example of such a situation during a Canadian peace-keeping mission is a 1993 incident in Somalia during which a 16 year-old Somali boy, who had been attempting to steal supplies, died at the hands of Canadian peacekeepers while in their custody. The “after-shocks” of this incident were felt and, nearly a decade later, continue to reverberate at all levels of the Canadian Forces.

Complicating the issue of working in an unfamiliar culture is the fact that the skills required by peacekeepers are incongruent with conventional military training, requiring as they do, attributes not normally assessed during the recruitment process (Schmidtchen, 1997a). This realization was the basis of Schmidtchen’s (1997a) thesis on the preparation of capable peacekeepers for the Australian Defence Force which was later published as an official publication of the Australian Defence Force (Australian Defence Force [ADF], 1999). One outcome of this document was the Cross Cultural Adaptability Scale (CCAS).

The CCAS is based on Schmidtchen’s (1997b) Cross Cultural Adaptability Model. In brief, the model comprises the following six subscales:

**Openness to Experience.** Peacekeepers must be open to new ideas because they are continually presented with a range of different cultures. These cultures may
be organizational (e.g., UN and other NGOs), military (e.g., peacekeepers from other nations or other services), or ethnically-based. The maintenance of consent, a fundamental principle of peace operations, is supported by the peacekeepers’ ability to recognise, adapt to, and integrate ideas and ways of doing that are different from their own.

**Attention to Interpersonal Relations.** People who are successful in cross-cultural environments tend to be skilled in their ability to recognise and attend to interpersonal relations. They are attentive to verbal and non-verbal communication cues and the context in which they occur; they are also sensitive to the impact of their behaviour on those with whom they interact, and can accurately communicate their intentions in a sensitive and meaningful way.

**Sense of Identity.** When entering cross-cultural situations, an individual’s beliefs and values may be fundamentally challenged. In a peace operation where exposure to different cultures can be intense, emotional, but inherently short term, it is important that service personnel have a high degree of self-awareness and maintain a strong sense of self-identity.

**Person-Organisation Goal Alignment.** Peace operations present service personnel with a range of new roles that stem from the strategic role of the organization in which they serve. In the context of peace operations, the UN is a third party mediator and monitor and this role must be reflected at every level of the force and in the actions of the individual. Consequently, it is important that peacekeepers’ values and beliefs be consonant with those of the organisation.

**Problem Solving.** Peace operations and cross-cultural environments consistently present service personnel with novel problems in a novel context. In these situations, the individual has limited ability to generalize from previous training and experience. In such a socially ambiguous environment, the ability to identify problems, produce novel solutions, and learn from the experience is fundamental to successful performance.

**Cross Cultural Experience.** Previous experience in, or exposure to, cross cultural conditions or a propensity to participate in these types of conditions can provide the individual with skills, abilities, and attitudes that can be successfully transferred to peace operations environment.

Depending on the robustness of the scale, the CCAS could be used as a tool to form the basis of cross cultural ability skills training to be provided to peacekeepers during pre-deployment preparation. It could also be used as a tool for selecting suitable individuals for peacekeeping operations.

Recognizing that all five TTCP countries contribute military forces to international peace operations, TTCP HUM TP3 agreed in 1998 to examine the potential for the application of the CCAS to the selection of peacekeepers. A first stage of such an examination was the confirmation of the
robustness of the psychometric properties of the questionnaire and its factor structure when administered to samples from the various TTCP countries. Ultimately, only Canada and Australia contributed test data to the study.

Method

Participants

Two sub-samples of military personnel, one Canadian (n = 751) and one Australian (n = 563), completed the CCAS. The Canadian responses (57.2% of the total sample) were extracted from a larger set of data, part of a survey that was administered in the spring of 1998. Of the 1314 total participants, 1006 (79.3%) were men and 263 (20.7%) were women (45 participants did not indicate their sex). In addition to differences in sex and country of service, the participants represented a wide range of ages, ranks, disciplines or commands, previous peacekeeping experiences, and previous cross-cultural experience of both samples.

Measure

Background Information. Participants indicated their age, sex, military occupation, rank, years of service, and country for which they served. Participants also indicated the number of countries that they had visited in the past five years.

Cross-Cultural Adaptability Scale (CCAS). The CCAS is a 53-item scale comprised of the following six subscales: Openness to experience (10 items); Attention to interpersonal relations (8 items); Sense of identity (10 items); Person-organisation goal alignment (10 items); Problem solving (5 items); and, Cross-cultural experience (10 items). With the exception of one item, participants were asked to rate their response to each item on a Likert-type scale ranging from one (strongly agree) to six (strongly disagree). One item (Have you visited other countries in the past five years?) was a dichotomous ‘yes/no’ question.

Because the questionnaire was originally developed for use with Australian military personnel, it was necessary to modify a number of the background data items before it could be administered in a Canadian context; these modifications are not considered to have affected the psychometric properties of the test. The only major difference in the administration of CCAS to both samples is that, in the Canadian administration, the questionnaire formed part of a larger study.

Results

Psychometric Properties of the CCAS

Internal consistency. To assess the internal consistency of the CCAS in the two national samples as well as the combined sample, Cronbach’s Alpha (?) was calculated for each of Schmidtchen’s six subscales and for the overall scale; the results of these calculations are presented by national sample in Table 1. The internal consistency
reliabilities for all of the subscales, with the exception of the Sense of identity, were acceptable (Guion, 1998).
Table 1

**Internal Consistency Reliabilities (Cronbach’s ? ) for the Six Subscales**

<table>
<thead>
<tr>
<th>Factors</th>
<th>Cronbach’s ?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Canada</td>
</tr>
<tr>
<td>Openness to Experience</td>
<td>.81</td>
</tr>
<tr>
<td>Attention to Interpersonal Relations</td>
<td>.86</td>
</tr>
<tr>
<td>Sense of Identity</td>
<td>.66</td>
</tr>
<tr>
<td>Person-Organisation Goal Alignment</td>
<td>.81</td>
</tr>
<tr>
<td>Problem Solving</td>
<td>.80</td>
</tr>
<tr>
<td>Cross-Cultural Experience</td>
<td>.81</td>
</tr>
<tr>
<td>Overall</td>
<td>.94</td>
</tr>
</tbody>
</table>

The inter-variable correlation matrix between the CCAS subscales and the total score was examined. Results are presented in Table 2. Inter-variable correlations for the six subscales ranged from \( r = .51 \) (\( N = 964 \)) and \( r = .68 \) (\( N = 1201 \)), suggesting that the subscales are not completely independent concepts. The CCAS subscales correlated highly with the total score ranging from \( r = .76 \) (\( N = 844 \)) to \( r = .83 \) (\( N = 844 \)). Although it is normal for the subscales of a test to correlate significantly with the total scores, these high correlations suggest that a single overall CCAS score could replace each of the six subscales.

Table 2

**Correlations Between CCAS Subscales and Total Score**

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Openness to Experience</td>
<td>.58*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Attention to Interpersonal Relations</td>
<td></td>
<td>.58*</td>
<td>.68*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Sense of Identity</td>
<td></td>
<td></td>
<td>.58*</td>
<td>.68*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Person-Organisation Goal Alignment</td>
<td></td>
<td></td>
<td></td>
<td>.52*</td>
<td>.54*</td>
<td>.52*</td>
</tr>
<tr>
<td>5. Problem Solving</td>
<td></td>
<td></td>
<td></td>
<td>.53*</td>
<td>.57*</td>
<td>.62*</td>
</tr>
<tr>
<td>6. Cross-Cultural Experience</td>
<td></td>
<td></td>
<td></td>
<td>.63*</td>
<td>.61*</td>
<td>.56*</td>
</tr>
<tr>
<td>7. Total Score</td>
<td>.81*</td>
<td>.82*</td>
<td>.83*</td>
<td>.76*</td>
<td>.76*</td>
<td>.82*</td>
</tr>
</tbody>
</table>

Note: * = \( p < 0.01 \)

**Comparing the Canadian and Australian Samples**

The comparability of the demographic characteristics of the two samples, was assessed through a multivariate analysis of variance (MANOVA); the following five dependent variables were examined: sex, age, rank, number of peacekeeping tours, and whether or not the respondent had visited other countries in the past five years. The results of the MANOVA (Wilks’ Lambda = 0.50, \( F(5, 1208) = 6462.97, p < .000 \)) indicated the presence of significant multivariate differences; subsequent, significant univariate F-ratios were obtained for all five dependent (demographic) variables (Table 3).
Additionally, Chi-square tests indicated significant differences between Canada and Australia in the distribution of ranks ($\chi^2_{(12,1262)} = 231.88, p < .000$) and types of command ($\chi^2_{(5,1280)} = 284.0588, p < .000$). Consequently, it was determined that separate, parallel analyses were necessary for each of the two nationalities.

Table 3

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>DF</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>9.83</td>
<td>1</td>
<td>9.83</td>
<td>62.00</td>
<td>.00</td>
</tr>
<tr>
<td>Age</td>
<td>251.11</td>
<td>1</td>
<td>251.11</td>
<td>198.68</td>
<td>.00</td>
</tr>
<tr>
<td>Rank</td>
<td>328.01</td>
<td>1</td>
<td>328.01</td>
<td>30.43</td>
<td>.00</td>
</tr>
<tr>
<td>Number of Peacekeeping Tours</td>
<td>178.06</td>
<td>1</td>
<td>178.06</td>
<td>200.50</td>
<td>.00</td>
</tr>
<tr>
<td>Countries Visited in the Past Five Years</td>
<td>40.65</td>
<td>1</td>
<td>40.65</td>
<td>199.14</td>
<td>.00</td>
</tr>
</tbody>
</table>

Factor Analysis of the Canadian Sample

Exploratory Factor Analysis (EFA) was conducted using principal components analysis with direct oblimin rotation. Component extraction technique was used to reduce the large number of variables measured by the CCAS to a small number of components (Tabachnick & Fidell, 2001, p. 612). Direct oblimin (oblique) rotation was used because the subscales of the CCAS correlated with each other (Tabachnick & Fidell, 2001, p. 616).

A free EFA was conducted initially to determine the number of factors. The decision was made to consider factor loadings above .4 significant. Results of the analysis indicated the presence of 12 factors with eigenvalues greater than one. However, the scree plot suggested that there were only five factors. Based on the scree plot, the decision was made to conduct a five-factor analysis.

The five factors accounted for 42.5% of the variance observed, which was acceptable (Table 4). These factors are interpreted below:

Factor 1, which contained items related to interpersonal relations, sense of identity, and problem solving, accounted for 25.5% of the total variance.

Factor 2, composed primarily of items related to openness to experience, accounted for 5.3% of the total variance.

Factor 3, which was composed mostly of items related to organisational goals, accounted for 4.5% of the total variance.

Factor 4, which contained items related to personal goals, accounted for 4.1% of the total variance.
Factor 5, which consisted of items related to interpersonal relations, accounted for 3.2% of the total variance.
Table 4

Rotated Factor Loading for Exploratory Factor Analysis of the CCAS – Canadian Sample

<table>
<thead>
<tr>
<th>Item</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>15. I try to understand other peoples’ thoughts and feelings when I talk to them</td>
<td>.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. I consider the impact my actions have on others</td>
<td></td>
<td>.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. I have personal standards of behaviour that I try to maintain</td>
<td>.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. I have a set of personal guidelines I use to decide what is right and wrong</td>
<td>.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42. I am a practical person</td>
<td>.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43. I enjoy adapting my skills to solve new problems</td>
<td>.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40. When things are slow I look for work</td>
<td>.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45. I can learn a lot from working with people from different backgrounds than me</td>
<td>.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24. I am sensitive to the needs of others</td>
<td>.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28. I am a good listener</td>
<td>.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44. I have a definite interest in learning about the local population</td>
<td>.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39. I can identify problems and develop innovative solutions</td>
<td>.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. I like taking risks</td>
<td>-.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. I like working in situations with no clear solution</td>
<td>-.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. I like being in unfamiliar situations</td>
<td>-.72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. I enjoy situations that require crisis management</td>
<td>-.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48. I feel comfortable in new situations</td>
<td>-.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. I can do my job even when things are not clear</td>
<td>-.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27. I deal well with stressful situations</td>
<td>-.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41. I can make critical decisions on the spur of the moment</td>
<td>-.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. I can operate where there are few rules.</td>
<td>-.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30. My country should continue to support peace operations</td>
<td>-.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29. The role of the UN/NATO is important</td>
<td>-.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32. My country should assist nations that need help</td>
<td>-.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31. I am a volunteer for overseas deployment</td>
<td>-.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49. There is a lot to be learned from working with peacekeepers from other nations</td>
<td>-.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. I like to travel</td>
<td>-.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52. It is important to learn as much of the local language as possible</td>
<td>-.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25. I like to have clearly stated tasks to achieve</td>
<td>.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. I like to know what is expected of me in advance</td>
<td>.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33. My main motivation for deploying overseas is financial reward</td>
<td>.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36. Should be greater recognition for those deployed</td>
<td>.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35. Gaining cooperation of unit is important on peace operations</td>
<td>.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I make friends easy</td>
<td>-.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. I like being around other people</td>
<td>-.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23. I have a good sense of humour</td>
<td>-.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. I am an outgoing person</td>
<td>-.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. I can make myself understood in most situations</td>
<td>-.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standardized Reliability: .86 .83 .83 .53 .74

With a cut-off score of .4, 14 of the 52 items did not load on any factor. There were no cross-loadings. With the exception of factor 4, the internal consistency values for each factor were found to be satisfactory (Factor 1: \( \alpha = .86 \); Factor 2: \( \alpha = .84 \); Factor 3: \( \alpha = \)
.83; factor 4: \( \gamma = .53 \); Factor 5: \( \gamma = .74 \). Deleting items from factor 4 would have had a negligible effect on the reliability, increasing it from .53 to .55.

**Factor Analysis of the Australian Sample**

An EFA was conducted on the Australian sample using principal components analysis with direct oblimin rotation. A free EFA was conducted to determine the number of factors: again, only items with factor loadings above .4 were considered significant. As with the Canadian sample, while results of the analysis indicated the presence of 12 factors with eigenvalues greater than one, the scree plot suggested that there were only five factors. Based on the scree plot, the decision was made to conduct a five-factor analysis.

The five factors accounted for 43.3% of the variance observed, which was acceptable (Table 5). These factors are interpreted below:

Factor 1, which contained items related to interpersonal relations and sense of identity, accounted for 27.2% of the total variance.

Factor 2, composed primarily of items related to openness to experience, accounted for 5.1% of the total variance.

Factor 3, which was composed mostly of items related to organisational goals and cross-cultural experience, accounted for 4.2% of the total variance.

Factor 4, which contain items related to personal goals, accounted for 3.6% of the total variance.

Factor 5, which contained items related to problem solving, accounted for 3.1% of the total variance.

With a cut off score of .4, 10 of the 52 items did not load on any factor. There were no cross loadings. With the exception of the fourth factor, the internal consistency values for each factor were found to be satisfactory (Factor 1: \( \gamma = .88 \); Factor 2: \( \gamma = .82 \); Factor 3: \( \gamma = .88 \); Factor 4: \( \gamma = .42 \); Factor 5: \( \gamma = .74 \)). Deleting items from factor 4 would have had a negligible effect on the reliability, increasing it from .42 to .43.

**DISCUSSION**

Although the CCAS was theoretically based on 6 factors, the results of independent factor analyses of the Canadian and Australian samples point to a five-factor structure. There are some differences between the two emergent factor structures, however, and the source of those differences is not readily discernible. They may be attributed to various factors such as level of peacekeeping experience (more than half of Canadian respondents (52.6%) indicated that they have been on one or more peacekeeping tours, while only 13.9% of Australian respondents have done so), experience in different countries (the majority of Canadian respondents (72.7%) have
visited other countries in past five years, while only 37.7% of Australian respondents have done so. There were also significant differences between both nations in terms of distribution of ranks.

Table 5
Rotated Factor Loading for Exploratory Factor Analysis of the CCAS – Australian Sample

<table>
<thead>
<tr>
<th>Item</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>24. I am sensitive to the needs of others</td>
<td>.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I make friends easy</td>
<td>.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. I like being around other people</td>
<td>.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. I am an outgoing person</td>
<td>.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. I try to understand other peoples’ thoughts and feelings when I talk to them</td>
<td>.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23. I have a good sense of humour</td>
<td>.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28. I am a good listener</td>
<td>.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. I am tolerant of other peoples’ attitudes and behaviours</td>
<td>.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. I am open-minded</td>
<td>.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. I am a confident person</td>
<td>.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. I consider the impact my actions have on others</td>
<td>.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. I like working in situations with no clear solution</td>
<td>.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. I like being in unfamiliar situations</td>
<td>.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. I like taking risks</td>
<td>.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. I enjoy situations that require crisis management</td>
<td>.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. I can operate where there are few rules.</td>
<td>.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27. I deal well with stressful situations</td>
<td>.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48. I feel comfortable in new situations</td>
<td>.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. I like to try new things</td>
<td>.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49. There is a lot to be learned from working with peacekeepers from other nations</td>
<td>.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31. I am a volunteer for overseas deployment</td>
<td>.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30. My country should continue to support peace operations</td>
<td>.72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32. My country should assist nations that need help</td>
<td>.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52. It is important to learn as much of the local language as possible</td>
<td>.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44. I have a definite interest in learning about the local population</td>
<td>.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45. I can learn a lot from working with people from different backgrounds than me</td>
<td>.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35. Gaining cooperation of unit is important on peace operations</td>
<td>.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29. The role of the UN/NATO is important</td>
<td>.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50. I enjoy talking to people who are different than me</td>
<td>.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34. I feel comfortable with the objectives of this deployment</td>
<td>.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. I like to travel</td>
<td>.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33. My main motivation for deploying overseas is financial reward</td>
<td>.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. I like to know what is expected of me in advance</td>
<td>.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25. I like to have clearly stated tasks to achieve</td>
<td>.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26. I am quick to judge other peoples’ character</td>
<td>.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41. I can make critical decisions on the spur of the moment</td>
<td>.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. I have a set of personal guidelines I use to decide what is right and wrong</td>
<td>.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39. I can identify problems and develop innovative solutions</td>
<td>.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42. I am a practical person</td>
<td>.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43. I enjoy adapting my skills to solve new problems</td>
<td>.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37. I regularly keep abreast of world news</td>
<td>.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Although it had lower factor loadings, the five-factor structure derived from the Australian sample was clearer than the Canadian sample. Additionally, with less missing data, skewness, kurtosis, and outliers, the Australian data set was the “cleaner” of the two. For these reasons, it was decided to use the five-factor solution derived from the Australian sample as the “tentative” solution. The five factors are thus interpreted as:

- Interpersonal relations/Sense of identity;
- Openness to experience;
- Organisational goals/Cross-cultural experience;
- Personal goals; and
- Problem solving.

In the course of the EFA of the Australian data, 10 of the 52 items (items 4, 5, 7, 17, 18, 36, 38, 40, 46, and 51) did not load significantly on any factors, which suggests that they should be dropped from the scale. It is of note that five of these items did not load significantly in the Canadian study (5, 17, 38, 46 and 51).

Canada’s peacekeeping experiences in Somalia have demonstrated the importance of selecting and training suitable personnel for peace operations. Although it is still under development, it is believed that the CCAS could one day form the basis of cross cultural adaptability skills training during peacekeeper pre-deployment training. If it can be shown that the CCAS can predict successful performance in peace operations, it could also see employment as a selection tool to identify suitable peacekeepers. Considerably more work remains, however, before this goal can be achieved. It is recommended that future work focus on revision of the CCAS and the conduct of field studies that focus on verifying the factor structure of the CCAS and determining the predictive validity of the scale. The field studies should include pre and post testing and comparison against independent performance criteria.

REFERENCES


Redeveloping the selection process for the Royal Australian Air Force Officer applicants

Jennifer Wheeler, M.Sc.
Psychology Research and Technology Group
Australian Department of Defence
Jennifer.Wheeler@cbr.defence.gov.au

Abstract

The Officer Interview Board (OIB) is the final stage in the selection process for applicants applying for Officer level positions in the Royal Australian Air Force (RAAF). Concerns were raised about the validity of the OIB in the selection of Officer applicants, including; limited standardisation between OIBs, high rejection rate of applicants, and the potential for the selection criteria assessed at the OIB to have become outdated. This study examined the OIB, in the context of the entire RAAF Officer selection system, to identify areas that could be refined and/or redeveloped. The study aimed to develop a selection system that was current, streamlined, standardised and a valid predictor of future training and job performance. A job analysis was conducted to determine the essential attributes of entry-level RAAF Officers, and subsequently identified nine attributes deemed essential to assess in the selection system. An examination of the RAAF Officer selection system revealed that a number of the essential attributes were not being assessed. In addition, it was maintained that some of the essential attributes could be assessed through assessment techniques other than an interview format. It was proposed that the OIB be redeveloped into an assessment-centre style Selection Board to assess the essential attributes of RAAF Officer applicants. This paper describes the development and final design of the RAAF Officer Selection Board. Future directions for the validation of the Selection Board will also be discussed.

Introduction

Selection into the Royal Australian Air Force (RAAF) was based, in part, on performance in a selection interview (Officer Interview Board). The interview was the final stage in the selection process for applicants applying for Officer level positions. To reach the selection interview, applicants must have successfully met the requirements of a medical assessment, psychological and cognitive testing, and interviews with a recruitment Officer and Psychologist.

In September 2001, concerns were raised by the Director of Personnel Officers - Air Force (DPO-AF) about the validity of the RAAF Officer Interview Board (OIB), in which applicants, who had met cognitive, academic and psychological requirements, were assessed by a selection interview. The concerns raised included:

a. High rejection rate of applicants at OIBs;

b. Limited standardisation between OIBs; and
c. Potential for the selection criteria assessed at the OIB to have become outdated or lacking in some areas.

A study was initiated to investigate the validity of the OIB, and in turn, identify areas that could be refined and/or developed into a new selection process. The aim of the study was to ensure that the RAAF Officer selection system was current, streamlined, standardised and a valid predictor of future job training and performance. This paper reports the development and final design of the RAAF Officer Selection Board in the context of the entire RAAF Officer selection system. Future directions for the validation of the new Selection Board is also discussed.

Background

RAAF Officer Interview Board (OIB)

The former selection interview (OIB) for applicants applying for Officer level positions in the RAAF was largely unstructured in nature. While the interview panel were given some direction as to how to conduct the interview (including the selection criteria), there existed, for example, no standard list of interview questions or standardised scoring system to arrive at a total effectiveness (hirability) rating. Recent research has supported the enhanced reliability (consistency) and validity (predicting future job training and performance) of structured interviews over unstructured interviews (for example, Campion, Palmer & Campion, 1998; Huffcutt & Arthur, 1994; McDaniel, Whetzel, Schmidt & Mauer, 1994; Wiesner & Cronshaw, 1988). This research had important implications for the former OIB format. An interview with limited structure may result in differing interpretations of standards and performance between boards. Different interpretations and standards have significant implications for the ‘top-down’ selection method used by DPO-AF to select applicants for RAAF Officer training positions.

Another concern raised in relation to the OIB was that the selection criteria assessed during the interview may not have been a current and comprehensive assessment of the essential characteristics of an entry-level RAAF Officer. A selection system that does not directly assess the essential characteristics of a respective occupation may result in inappropriate appointments and subsequent wastage (time and resources) for a number of stakeholders. The essential characteristics of a specific occupation can be established through conducting a job analysis, which in turn, is a component of structure in a selection interview (Campion et al., 1998).

It was proposed that the OIB incorporate components of a structured interview (for example, standardised interview questions based on a job analysis) to enhance its reliability and validity, and hence, its usefulness in decision making and prediction. Furthermore, structural components of an interview facilitate the assessment of applicants through reducing the memory requirements of the interviewers and standardising the sample of behaviours being judged by the interview board (Conway, Jako & Goodman, 1995; Dipboye & Gaugler, 1993).
Selection Criteria Development Questionnaire

To develop a job-related selection criteria for RAAF Officers, the first stage of the study was to conduct a job analysis, namely the Selection Criteria Development Questionnaire (SCDQ, see Steele & Marr, 2002 for review). The SCDQ utilises subject matter experts (SMEs), such as experienced specialisation members, to determine the essential attributes (abilities, knowledge and personal qualities) required for successful training and job performance. A selection system that is clearly linked to the requirements of a job assists in demonstrating fair and legally defensible selection decisions. Furthermore, a job-related selection system will provide a number of gains for the recruitment process, through providing more accurate job-person matches, resulting in a reduction in failure rates and attrition, and also supporting the Australia Defence Organisation’s goal of the ‘right person for the right job’ (Steele & Close, 2002).

The SCDQ revealed nine essential characteristics of entry-level RAAF Officers. The findings were based on the responses of 124 SMEs (response rate of 65%), covering all RAAF specialisations. The essential characteristics of entry-level RAAF Officers are presented in Table 1.

**Table 1. Attributes Assessed as Essential for RAAF Officer by 75% of Subject Matter Experts**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Importance of Attribute</th>
<th>Level of Attribute</th>
<th>% of SMEs identifying attribute as essential</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Standard Deviation</td>
<td>Mean</td>
</tr>
<tr>
<td>1. Oral Comprehension</td>
<td>2.95</td>
<td>0.23</td>
<td>5.98</td>
</tr>
<tr>
<td>2. Written Comprehension</td>
<td>2.90</td>
<td>0.30</td>
<td>5.59</td>
</tr>
<tr>
<td>3. General Cognitive Ability</td>
<td>2.88</td>
<td>0.33</td>
<td>5.51</td>
</tr>
<tr>
<td>4. Oral Expression</td>
<td>2.86</td>
<td>0.34</td>
<td>5.24</td>
</tr>
<tr>
<td>5. Security Mindedness</td>
<td>2.85</td>
<td>0.38</td>
<td>5.74</td>
</tr>
<tr>
<td>6. Conscientiousness</td>
<td>2.80</td>
<td>0.40</td>
<td>4.66</td>
</tr>
<tr>
<td>7. Written Expression</td>
<td>2.78</td>
<td>0.42</td>
<td>5.02</td>
</tr>
<tr>
<td>8. Speech clarity</td>
<td>2.77</td>
<td>0.42</td>
<td>5.29</td>
</tr>
<tr>
<td>9. General Leadership Ability</td>
<td>2.76</td>
<td>0.43</td>
<td>4.89</td>
</tr>
</tbody>
</table>

*Note: Importance of Attribute ranges from 1 (not required) to 3 (essential characteristic).*
*Note: Level of Attribute ranges from 1 (low level) to 7 (very high level).*
*Note: Gatewood and Feild (1994) recommend using a cut-off of 75% agreement to determine inclusion of an ability as an essential characteristic.*

10 The SCDQ has been successfully used in the analysis of a number of Australian Defence Force occupations.
Review of the essential characteristics in the RAAF Officer selection system

Based on the findings of the SCDQ, the next stage of the project involved reviewing the assessment of the essential characteristics in the RAAF Officer selection system. An analysis of the RAAF Officer selection system revealed that a number of the essential characteristics were not being assessed. In addition, it was proposed that some of the essential characteristics (examined in the OIB) could be assessed more directly through assessment techniques other than an interview format. The essential characteristics, OIB and the proposed assessment methods to assess the respective characteristics are presented in Table 2.

Table 2. Essential RAAF Officer characteristics (and definitions), OIB and proposed assessment methods to assess the respective characteristic.

<table>
<thead>
<tr>
<th>Essential Characteristic</th>
<th>Definition</th>
<th>OIB Assessment Method</th>
<th>Proposed Assessment Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral Comprehension</td>
<td>The ability to understand spoken English words and sentences.</td>
<td>Not assessed</td>
<td>Oral Comprehension Test</td>
</tr>
<tr>
<td>Written Comprehension</td>
<td>The ability to understand written sentences and paragraphs.</td>
<td>Indirectly via AGC</td>
<td>Indirectly via AGC</td>
</tr>
<tr>
<td>General Cognitive Ability</td>
<td>The primary factor among the many factors comprising intellectual capacity.</td>
<td>AGC</td>
<td>AGC Structured Interview</td>
</tr>
<tr>
<td>Oral Expression</td>
<td>The ability to orally communicate information and the meaning of the words to other people.</td>
<td>Interview</td>
<td>Oral Presentation Structured Interview</td>
</tr>
<tr>
<td>Security Mindedness</td>
<td>The appreciation for the need for security and to put into practice security guidelines.</td>
<td>Not assessed</td>
<td>Oral Presentation</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>--------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>A general tendency to work hard and be loyal, to give a full day’s work and to do one’s best to perform well.</td>
<td>Not assessed</td>
<td>Trait Self-Description Inventory Structured Interview</td>
</tr>
<tr>
<td>Written Expression</td>
<td>Involves knowledge of grammar, the meaning of words and how to organise sentences and paragraphs.</td>
<td>Indirectly via Writing Skills exercise (essay)</td>
<td>Writing Skills exercise (essay)</td>
</tr>
<tr>
<td>Speech Clarity</td>
<td>The ability to communicate orally in a clear fashion that is understandable to a listener.</td>
<td>Indirectly via Interview</td>
<td>Oral Presentation Structured Interview</td>
</tr>
<tr>
<td>General Leadership Ability</td>
<td>The tendency to initiate action, to take charge of situations or groups, and to influence or motivate behaviour or thinking of other persons or groups of people.</td>
<td>Interview</td>
<td>Leadership Exercise Structured Interview</td>
</tr>
</tbody>
</table>

Based on the findings of the SCDQ and research conducted into the validity of different assessment techniques and methods, it was proposed that:

a. The Officer Interview Board (OIB) be redeveloped into a Selection Board (assessment centre) to assess the essential characteristics of entry-level RAAF Officers; and

b. The assessment of conscientiousness (via the Trait Self-Description Inventory) and written expression (via a structured writing skills exercise) be implemented at the recruitment centre stage.

Therefore, the selection system for RAAF Officers applicants comprises:

a. Recruitment centre assessment (psychological and cognitive testing);

b. Psychological report
This report presents the design of the new RAAF Officer Selection Board (ROSB), which replaces the former OIB as the final stage in the selection process for applicants applying for RAAF Officer positions.

**RAAF Officer Selection Board (ROSB)**

The ROSB is a structured combination of assessment techniques that are used to provide a wide-ranging assessment of each applicant. The ROSB consists of a one-day assessment centre\(^\text{11}\), comprising assessment tasks in the morning, followed by individual structured interviews. The assessment techniques in the ROSB are designed to be job related (assess the essential characteristics of RAAF Officers) and designed to show observable dimensions and behaviours. The structured nature of the ROSB allows objective and standardised assessment of RAAF Officer applicants both within and between Selection Boards.

**Assessment Tasks**

The assessment tasks employed during the morning of the ROSB are as follows:

- **Oral Comprehension Test (Test OC):** Test OC is designed to assess the ability to understand the meaning of spoken English words, phrases, sentences and paragraphs. Based on an oral passage delivered by tape (approximately six minutes in length), applicants’ understanding of the content contained in the passage is assessed by 15 multiple-choice questions;

- **Leadership Exercise:** The group leadership exercise is designed to assess applicants’ potential for, and natural leadership abilities. The leadership exercise (‘survival’ type exercise) assesses applicants’ leadership styles, behaviours and the ability to work as a member of a team; and

- **Oral Presentation:** Applicants are required to present prepared oral presentations of five minutes in duration to the selection board. The individual oral presentations assess applicants’ oral communication skills, namely oral communication and speech clarity. In addition, the content of the presentation is also assessed, specifically applicants’ knowledge of the RAAF Officer position and security mindedness.

\(^{11}\) Designed to assess six applicants per board day.
Structured Interview

Following the assessment tasks, applicants are assessed by a structured interview. The interview panel consists of a Board President, Psychologist and Specialist Officer. The interview assesses a number of the essential characteristics determined from the SCDQ (see Table 2), in addition to ‘suitability for the specialisation’, ‘social adjustment’ and ‘ability and aptitude’. The following structural components have been implemented to increase the reliability and validity of the interview:

a. Interview questions are based on a job analysis (SCDQ);

b. Standard list of core interview questions to assess each criteria;

c. Behaviourally anchored rating scales; and

d. A standardised scoring system to arrive at a total effectiveness (hirability) rating.

Validation of the ROSB

A 12-month trial of the ROSB was approved by the DG-Personnel, with the new format being implemented as part of the RAAF Officer selection system in September 2002. The initial predictive validation study (approximately 12 months from the implementation date) will focus on the validity of the ROSB in predicting performance in job training at the Officer Training School (Direct Entry Boards) and Australian Defence Force Academy (ADFA Boards). Further research studies will continue to analyse the predictive validity of the ROSB in predicting job training, with subsequent long-term studies analysing the ROSB’s ability to predict future job performance (for example, promotions).

References


Audit of the United States Air Force Promotion Testing Program

Paul P. Stanley II
Chief, Test Development Flight
Air Force Occupational Measurement Squadron
Randolph Air Force Base
San Antonio, Texas

Written tests have been an integral part of the Air Force enlisted promotion program since 1970. The Air Force Occupational Measurement Squadron (AFOMS) develops these tests using procedures designed to be congruent with the highest standards of the civilian testing community. Because test results are key to Air Force enlisted promotion decisions, the Headquarters Air Force personnel directorate determined that a review of the process should be conducted, and that the review should be conducted by a respected authority outside of the Air Force itself. The organization selected to conduct the audit was Chauncey Group International (CGI), a subsidiary of the Educational Testing Service. The CGI audit report gave the Air Force testing program an overall rating of Outstanding.

BACKGROUND

Each year, AFOMS’s Test Development Flight publishes approximately 320 new tests, which are used in the enlisted promotion decision process. The tests are developed by specially selected senior noncommissioned officer subject-matter experts (SMEs) under the supervision of AFOMS behavioral scientists. The SMEs provide the technical expertise, using task data gathered by the Occupational Analysis Flight of AFOMS, while military and civilian psychologists from AFOMS provide the psychometric expertise.

Standard test development procedures are extensively described in the AFOMS Handbook for the Development of the SKT and Associated Tests. The Handbook has been carefully refined over the years to ensure consistency with professional guidance in the field, including the Society for Industrial and Organizational Psychology’s Principles for the Validation and Use of Personnel Selection Procedures and the US government’s Uniform Guidelines on Employee Selection Procedures.

PURPOSE OF AUDIT

Because test results play such a key role in Air Force enlisted promotion selections, the Headquarters Air Force personnel directorate determined, early in 2001, that a thorough review of the process should be conducted. This concern proved quite timely when Air Force Times, a civilian newspaper aimed at members of the US Air Force, took the editorial stance that written tests appeared to be unfair to members of minority groups. Although the newspaper’s own statistics failed to substantiate the view that tests resulted in unfair treatment of minorities, its provocative headlines, such as “Left Behind” and “Scores Don’t Say It All,” were a cause for concern at the Air Staff.

To ensure objectivity, Air Force managers sought an independent agency to audit the enlisted testing program. Having identified the need for an organization whose standing was recognized by the public at large and by members of the professional testing community, the Air Force chose the Chauncey Group International (CGI) to conduct the audit. The Chauncey Group website may be found at http://www.chauncey.com/. CGI is a subsidiary of the Educational Testing Service of Princeton, New Jersey. ETS is the company that produces the widely used and respected Graduate Record Examination (GRE) and the Scholastic Achievement Test (SAT).

AIR FORCE ENLISTED PROMOTIONS

Written tests play a key role in US Air Force enlisted promotions. Airmen compete for promotion to grades E-5, -6, and -7 under the Weighted Airman Promotion System (WAPS). They compete for promotion to grades E-8 and -9 under the Senior NCO Promotion Program (SNCOPP). Under
both WAPS and the SNCOPP, members are rank-ordered according to a single promotion score calculated based on individual objective promotion factors. The actual promotion score cut-off is calculated based on the overall percentage of eligibles who will be promoted, a number set by the Air Staff. Eligible airmen compete only against other airmen in the same specialty.

<table>
<thead>
<tr>
<th>Possible Points</th>
<th>WAPS</th>
<th>SNCOPP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specialty Knowledge Test (SKT)</td>
<td>100</td>
<td>----</td>
</tr>
<tr>
<td>Promotion Fitness Examination (PFE)</td>
<td>100</td>
<td>----</td>
</tr>
<tr>
<td>Enlisted Performance Ratings</td>
<td>135</td>
<td>135</td>
</tr>
<tr>
<td>Time in Grade</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>Time in Service</td>
<td>40</td>
<td>25</td>
</tr>
<tr>
<td>Awards &amp; Decorations</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Evaluation Board Score</td>
<td>----</td>
<td>450</td>
</tr>
<tr>
<td>USAF Supervisory Examination (USAFSE)</td>
<td>----</td>
<td>100</td>
</tr>
<tr>
<td>Total Possible Points</td>
<td>460</td>
<td>795</td>
</tr>
</tbody>
</table>

**Figure 1. Summary of WAPS and SNCOPP Promotion Factors**

**SKT.** The SKT is a 100-question multiple-choice test that measures knowledge shown to be important to performance in the member’s assigned specialty. Topics covered in SKTs are limited to those contained in the specialty training standard for the examinee’s career field. Test content is based on task-based products provided by the AFOMS’s Occupational Analysis Flight (AFOMS/OA). There are approximately 150 career fields, and there are generally two SKTs written for each: an E-5, for promotion to the rank of staff sergeant; and an E-6/7 SKT, for promotion to the ranks of technical and master sergeant. All study references used for SKT development are listed in the *WAPS Catalog*, which is made available to all examinees over the Internet.

**PFE.** The PFE is a 100-question multiple-choice test that measures military and supervisory knowledge required for successful performance in grades E-5, -6, and -7. The PFE is written entirely from Volume 1 of the *PFE/USAFSE Study Guide*, which is issued to every airman competing for promotion. Test content and study guide content are both based on occupational analysis data provided by AFOMS/OA. Four content-equivalent PFES are written for each grade.

**USAFSE.** The USAFSE is a 100-question multiple-choice test that measures military, supervisory, and managerial knowledge required for successful performance in grades E-8 and -9. The USAFSE is written entirely from Volumes 1 and 2 of the *PFE/USAFSE Study Guide*, which is issued to every airman competing for promotion. Test content and study guide content are both based on occupational analysis data provided by AFOMS/OA. Four content-equivalent USAFSEs are written for each grade.

**CHRONOLOGY OF EVENTS**

The audit was performed in three phases:

**Phase I: Planning and Preparation.** First, the respective points of contact from AFPC and AFOMS gathered all pertinent directives, including policy directives and procedural handbooks, and shipped them to CGI so that the auditors could begin comparing the Air Force testing program with the *ETS Standards of Quality and Fairness*, which may be found on the Internet at [http://www.ets.org/fairness/download.html](http://www.ets.org/fairness/download.html). The AFOMS point of contact took the additional step of preparing a comprehensive audit preparation report. The 60-page report, which was provided...
in its entirety to the CGI audit team, was based on the standards portion of the *ETS Standards of Quality and Fairness*. After each standard was written a description of how Air Force policies and procedures complied with that standard or, as was sometimes the case, how the ETS standard was not applicable in the Air Force setting. Figure 2 shows an illustrative example from the audit preparation report:

**ETS Standard 6.5:** Warn potential users to avoid likely uses of the assessment for which there is insufficient validity evidence.

No program can warn against (or even imagine) all the unsupported uses or interpretations that might be made of assessment results. However, experience may show that certain unsupported uses of the assessment results are likely. For example, licensing tests have been misused for job selection. If there is evidence that unsupported uses are likely (or are occurring), programs should warn users to avoid unsupported uses, and inform users of appropriate actions.

**How the Air Force complies with ETS Standard 6.5:** Air Force policy prohibits the use of promotion test results in any way other than to determine the promotion score of a qualified examinee competing for promotion:

“Organizations are prohibited from using any test or assessment (including commercial tests or assessments) not listed in AFIND 7 for any personnel action, no matter how small in scope or magnitude, without written approval or waiver by HQ USAF/DPF. No change in the manner in which an approved test is used may be made without approval of HQ USAF/DPF.”

(AF1 36-2605, para 1.2.1)

Figure 2. Excerpt from AFOMS Audit Preparation Report

**Phase II: Assessment of Policies and Procedures** The CGI auditors used the audit preparation report to prepare. By the time they arrived in San Antonio to meet personally with AFPC and AFOMS personnel, they had obviously done their own pre-audit preparation, as they arrived with an impressive understanding of the Air Force program, and with questions at the ready on issues that they did not understand, or thought might present a problem.

**Phase III: Assessment of Compliance.** Having found Air Force policies and procedures to be congruent with ETS Standards, the CGI auditors then examined 17 different tests (and all test-related documentation) and observed operations at five different test administration locations to ensure that Air Force personnel actually complied with the prescribed policies and procedures.

**Project Timelines**

September 2001: The Air Staff let the contract through standard channels.

October 2001: CGI met with Air Staff, AFPC, and AFOMS officials to plan audit.

November 2001: Air Force officials provided CGI with test-related directives and other guidance.

January 2002: CGI team leader visited AFOMS and AFPC to fine-tune audit plan.

February 2002: CGI audit team -- a psychologist, a psychometrician, and a generalist -- visited AFPC and AFOMS, as well as the three test control officers (TCOs) in the San Antonio area.
March 2002: CGI auditors visited several additional TCOs in the Washington DC area to complete their observations.

May 2002: CGI auditors reviewed 17 AFOMS tests (PFE, USAFSE, and 15 SKTs) and all of the statistical and validity documentation maintained for each. During the same visit, they conducted interviews with airmen who were recent examinees.

September 2002: CGI team chief delivered final outbrief to Air Force senior leadership.

RESULTS

As shown in the extract below, the final report gave the Air Force testing program an overall rating of Outstanding (Possible ratings: Outstanding, Excellent, Satisfactory, Unsatisfactory). Every applicable ETS standard was met or exceeded.

“The Chauncey Group International (a subsidiary of Educational Testing Service (ETS), Princeton, NJ) conducted an audit of the United States Air Force development and administration procedures and operations with regard to the USAF Supervisory Examination (USAFSE), Promotion Fitness Examination (PFE), and Specialty Knowledge Test (SKT) examination program. The audit was conducted from February through May 2002.

“Methodology

“The format used for conducting the audit consisted of observing personnel, reviewing documents, and discussing procedures with Air Force Occupational Measurement Squadron (AFOMS), Air Force Personnel Center (AFPC) and Test Center Management (TCM) staff and relating these observations to the ETS Standards for Quality and Fairness, a set of standards that provide a comprehensive list of critical features essential for quality testing programs. The standards are organized according to major features and conform to the process identified in the Chauncey Group’s proposal to the AFOMS. For purposes of this document, we have organized our report according to the standards applicable to the assessments including the USAFSE, PFE, and SKTs.

“Chauncey Group staff had unlimited access to the testing environment including all administration and secure areas. We were permitted access to all testing components and, in most cases; two independent observations were made for each area evaluated by the audit. Five test centers were visited, two in San Antonio, and three in the Washington, DC area as part of this audit. Program staff, Test Center examiners, and airmen were informed of the purpose of this audit and were all readily available to answer questions and to share their perspectives on the USAF examination programs audited.

“Findings

“Our findings are based on our review of available documents; conversations with AFOMS and AFPC personnel, test center examiners, airmen candidates, and five visits to different test centers.

“Findings are reported for each category and are scaled in the following order: Outstanding, Excellent, Satisfactory and Unsatisfactory. These ratings reflect the demonstrated compliance with the standards outlined in each category and therefore reflect the compliance of the whole program to industry standard. We rate the overall program as Outstanding. Our findings and recommendations are summarized below. No single category fell below Excellent; the
recommendations given are made to suggest a manner in which the finding might be resolved.”

Figure 3. Extract from Chauncey Group Audit Executive Summary
The ratings by individual category were as follows:

<table>
<thead>
<tr>
<th>ETS Standards</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer Service</td>
<td>Outstanding</td>
</tr>
<tr>
<td>Fairness</td>
<td>Excellent</td>
</tr>
<tr>
<td>Uses and Protection of Information</td>
<td>Outstanding</td>
</tr>
<tr>
<td>Validity</td>
<td>Outstanding</td>
</tr>
<tr>
<td>Assessment Development</td>
<td>Outstanding</td>
</tr>
<tr>
<td>Reliability</td>
<td>Outstanding</td>
</tr>
<tr>
<td>Cut Scores, Scaling, and Equating</td>
<td>Excellent</td>
</tr>
<tr>
<td>Score Reporting</td>
<td>Outstanding</td>
</tr>
<tr>
<td>Assessment use</td>
<td>Outstanding</td>
</tr>
<tr>
<td>Test Takers’ Rights and Responsibilities</td>
<td>Outstanding</td>
</tr>
</tbody>
</table>

**Overall Rating**  
Outstanding

In addition, the final report included several specific recommendations to consider, including such things as repeating an earlier customer satisfaction survey, performing a reading level analysis on the examinations, and using a more sophisticated statistic to assess differential item functioning.

**DISCUSSION**

AFOMS staff members expressed appreciation for the thorough and professional manner in which the CGI professionals conducted the audit. In particular, they commented on the audit team members’ evident appreciation for the practical side of test administration in addition to the already expected mastery of the academic aspects of testing. In the opinion of this author, any agency with a professionally developed testing program in place should not hesitate to request a similar review of their own program by CGI. If confidentiality is a concern, the contract can be written so as to preclude release without consent. (The Air Force audit final report was released in its entirety by the AFPC public affairs office.)

The final report of the audit supported all elements of the Air Force testing program. Before the final report had been written, the CGI team chief outbriefed the AFOMS staff, saying, “You have an outstanding program. You don’t need to change a thing.” Each of the recommendations in the final report will be formally evaluated by Air Force managers.

Considerable effort went into producing the Air Force program managers’ comprehensive audit preparation report, and the result was especially important to the efficient conduct of the audit. It was felt that Air Force promotion testing might be considerably different from other programs audited by CGI, and the report was aimed at highlighting those differences. Much of the report was incorporated word-for-word in the final report, with credit given to the Air Force writers.

Following the formal audit outbrief on September 25, 2002, the public affairs office prepared a news story for immediate release. The next day, an *Air Force Times* reporter conducted an interview with Air Force officials. The resulting article, which appears in the October 14, 2002, issue of *Air Force Times*, is shown at Appendix A.

**REFERENCES**


SAN ANTONIO - One of the most famous names in the testing business gives the Air Force's promotion testing system its top rating. At the same time, it recommends that test designers conduct a customer satisfaction survey to see how the troops in the field view the tests.

The Chauncey Group International, a subsidiary of Educational Testing Service of Princeton, N.J., conducted a top-to-bottom audit of the Weighted Airman Promotions System between February and May. Educational Testing Service is the company behind the Scholastic Achievement Test, the Graduate Record Examinations and other tests used by colleges and businesses. Chauncey gave the program an overall "Outstanding," its highest rating. Nine of 11 categories were rated as "Outstanding" and the remaining two as "Excellent", the second-highest rating.

"They brought in a team of psychologists and psychometricians and basically just went over our office with a fine-toothed comb,' said Paul 'Monty' Stanley, chief of the Test Development Flight of the Air Force Occupational Measurement Squadron at Randolph Air Force Base, Texas. Plans are already being made to design and conduct the customer survey, Stanley said. The last survey was an informal one in 1996.

"In '96, we sent a couple of folks to 10 or 11 bases to ask test takers some questions, to get a sense for how the customer felt, because one of our goals has always been to have a test that's not only valid and fair, but accepted by the force," Stanley said. "We learned that we can never stop educating people about the system. When people understand the system, there are fewer misconceptions and rumors, and thus greater acceptance."

Some airmen expressed concerns that one career field might have a harder test than another. But since airmen within each Air Force Specialty Code compete against only other airmen taking the same test, the relative difficulty among career fields is not a factor in promotions, Stanley said.

"The auditor team chief said to me, 'You have an outstanding program. You don't need to change a thing,'" Stanley said. "Which I took to be a full endorsement of the way that we do business."

Auditors focused on the Supervisory Examination, the Promotion Fitness Examination and 15 individual Specialty Knowledge Tests.

The tests are the cornerstone of the Weighted Airman Promotion System, usually making the difference between who does and does not get promoted.

Auditors suggested 16 possible changes, but most were in the arcane world of statistical analysis, different ways of looking at data.

"I think the nature of their underscores how strong this program is. It's very much fine-tuning things where you are really getting into the weeds," said Ken Schwartz, chief of the Test Management Section at the Air Force Personnel Center.

Although much of the testing world has been moving to automated tests, auditors said that probably would not be a good idea for the Air Force. "They recommend that we do not automate, that it makes more sense to stay with paper and pencil and continue to monitor and look at that area over future years," Schwartz said.

The audit, which cost about $100,000, was conducted using the "ETS Standards of Fairness and Quality," widely used in the testing industry. Auditors also visited five local test centers to observe how tests were administered and controlled. Auditors looked at all the documentation for tests, including records showing how each question can be tied to a specific task in a career field. They were very impressed with the way tests are developed using senior NCOs who are subject matter experts to write the actual questions, Stanley said.

"Obviously, we're very pleased that the program has gotten such an outstanding evaluation by the Chauncey Group," said Paul DiTullio, chief of Air Force Testing and Survey Policy at the Air Staff. "We felt confident that the program was being run well and that all our procedures were good. But it's always nice to have an outside, independent organization come in and do an assessment." The Test Development Flight has a Web site with information about tests at http://www.omsq.af.mil/OMD.
MENTAL HEALTH SUPPORT FOLLOWING CRITICAL INCIDENTS: THE AUSTRALIAN DEFENCE FORCE MODEL FOR ACUTE INTERVENTION

Captain Neanne Bennett
Australian Centre for Posttraumatic Mental Health
Austin & Repatriation Medical Centre,
P O Box 5444, Heidelberg West, Victoria, Australia 3081
neanne@unimelb.edu.au

ABSTRACT

This paper examines the effectiveness of early intervention, including the implementation of Critical Incident Stress Management (CISM) and psychological debriefing following potentially traumatic events and the relevance and implications of these practices on organisations such as the Australian Defence Force (ADF). Research into the effectiveness of these practices is limited and has produced mixed conclusions. Whilst there is some evidence to indicate that psychological debriefing provided to individuals post trauma has significant benefits, there is also substantial evidence to suggest that its effects may be short term and may contribute to longer term psychopathology. This paper reviews a number of studies examining the effectiveness of early intervention in civilian and military settings, considering the potential risks involved in providing psychological debriefing as well as alternative means of intervention that may be appropriate. Based upon empirical evidence and current worlds best practice, the ADF have developed a model for providing mental health support and intervention following critical incidents. The ADF’s Critical Incident Mental Health Support (CMS) model and its implementation will be discussed.

It has become recognised in many circles that mental health intervention should be provided to individuals exposed to traumatic events, however considerable debate has emerged over the effectiveness and implementation of such interventions. Early interventions, for the purposes of this paper, are generally classified as forms of primary prevention, (approaches that aim to reduce the incidence of new cases through intervention before the disorder occurs) or secondary prevention (approaches that aim to reduce the prevalence of disorder through early identification of problems with intervention before the disorder becomes severe). These interventions can range from low-key, practical assistance through to highly specialised psychological and pharmacological intervention models, such as Psychological Debriefing (PD) or Critical Incident Stress Debriefing (CISD). Some theorists state that the earlier the intervention is applied, the less risk of developing maladaptive cognitive and behavioural patterns (Neria & Solomon, 1999), however such claims are yet to be supported empirically. A fundamental point in the aftermath of trauma is to expect normal recovery and the expectation of clinically significant mental health problems occurring post-incident is inappropriate. Early intervention should be designed to promote normal and healthy adjustment processes and the purpose of this paper is to review the effectiveness of early intervention following psychological trauma and its application to the Australian Defence Force.
WHAT IS A TRAUMATIC EVENT

Although there is no definitive answer to what constitutes a traumatic event or critical incident, the Diagnostic and Statistical Manual (DSM) criteria for posttraumatic stress disorder (PTSD) and acute stress disorder (ASD) indicates exposure (either direct or indirect, such as hearing of the event) to a real or perceived threat affecting the physical integrity of the self or others may cause stress reactions in some individuals (American Psychiatric Association, 1994). More liberal definitions define a critical incident in terms of the psychological response rather than the event itself. Thus, any event that is likely to produce a strong emotional reaction would be defined as a “critical incident” (Mitchell & Everly, 1996; Miller, 1999). Thus support following a broad range of events that include unexpected death or serious physical injury (or perceived threats thereof) to personnel, their colleagues, or others with whom they have contact may be considered appropriate.

MODELS OF EARLY INTERVENTION

Psychological debriefing was originally designed to respond to individuals who experienced psychological trauma by providing a structure for establishing support and preparing individuals for recovery (Shalev, 2000; Litz, Gray, Bryant, & Adler, 2001). Conceptually the term “debriefing” is somewhat elusive and ill defined, as it refers to a heterogenous array of interventions including abjections, cognitive restructuring, self-diagnosis and education and can occur in a group or individual setting, as a single intervention or applied over several sessions (Neria & Solomon, 1999; Solomon, Neria & Witztum, 2000). The rationale and theory behind the process appears varied, thus support for the intervention appears best based upon empirical evidence, rather than theoretical constructs.

Modern approaches to psychological debriefing were developed predominately by Raphael (1986) and Mitchell (1983) to help rescue workers deal with the stress associated with their employment. Critical Incident Stress Management (CISM) was one such approach and was developed in the 1980s to assist emergency services personnel after exposure to traumatic or “critical” incidents and is the most commonly used method of psychological intervention. The model provided an opportunity for personnel to discuss their experiences in a rational, structured environment as well as to diffuse any emotional reactions (Mitchell, 1983, 1986).

CISM involves a range of crisis intervention and stress management services which may be offered to personnel at different time points following traumatic exposure and includes a psychological debriefing or critical incident stress debriefing (CISD) component (Mitchell & Everly, 1996; Miller, 1999; Mitchell & Everly, 2000; Everly, Flannery & Mitchell, 2000). However it is emphasised that CISD is only one part of CISM and that it should occur conjunctly with other procedures to form a comprehensive approach to the management of traumatic reactions. The most widely used form of debriefing is often referred to as the “Mitchell Model” (Mitchell & Everly, 1996; Everly & Mitchell, 2000), which consists of seven stages involving a review of thoughts,
emotions and symptoms related to the event. The stages of the model are outlined in Table 1.

Table 1: Stages commonly used in the Mitchell model of debriefing

<table>
<thead>
<tr>
<th>Stage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction and guidelines for participation</td>
</tr>
<tr>
<td>2</td>
<td>Discussion of the relevant facts</td>
</tr>
<tr>
<td>3</td>
<td>Discussion of thoughts</td>
</tr>
<tr>
<td>4</td>
<td>Discussion of reactions and emotions</td>
</tr>
<tr>
<td>5</td>
<td>Discussion of emergent symptoms</td>
</tr>
<tr>
<td>6</td>
<td>Education about responses and coping strategies</td>
</tr>
<tr>
<td>7</td>
<td>Re-entry (including summarisation, discussion of available resources, etc)</td>
</tr>
</tbody>
</table>

CISD is usually a group intervention that includes education on reactions and symptoms commonly experienced after a traumatic event. Essentially, it allows an opportunity to process and provide information on the traumatic experience and is often viewed as an opportunity to share responses to situations rather than a clinical intervention. In general, personnel are encouraged to attend a brief regardless of their level of symptomatology, with the underlying assumption being that all individuals will experience a reaction at some level. Although the characteristics of those conducting the debriefing varies, with mental health, allied health, and peer support personnel being used in different settings, there is general agreement that those providing debriefing should be trained and that standard protocols should be implemented and carefully followed. (It should be noted, however, that there is no evidence at this stage to suggest that training increases efficacy).

Whilst the Mitchell Model is considered the most widely used debriefing model worldwide, another popular model is that proposed by Raphael (1986). This model has similarities to the Mitchell Model and was developed to help teams of rescue workers involved with the Granville rail disaster (Raphael, Singh, Bradbury, & Lambert, 1983), in which formal group sessions were used to provide a cognitive structure and an emotional release for the workers. The Raphael model places an emphasis on workers being able to express their feelings and reactions to the event, discussion of symptoms they may experience, such as intrusive thoughts or nightmares and discussion of their relationships with family members and colleagues.

Over recent years, the Raphael and Mitchell models have been used successfully worldwide, however in attempts to refine its clinical adaptability and suitability amongst various populations, some variations from the original model have developed, however all tend to have retained the basic rationale of the stress debriefing approach (Dyregrov, 1989, 1998; Miller, 1999). For example, variations on the Mitchell Model have been developed for law enforcement personnel, disaster workers, workers at risk of workplace
violence as well as health professionals involved in working with trauma victims. A review of these models can be found in Miller (1999).

Given its popular use amongst the emergency services, and the similarities existing between many emergency situations and military activities, the Mitchell Model has also been adopted into many military environments, including the Australian Defence Force (Department of Defence, 1999).

**EFFICACY OF PSYCHOLOGICAL DEBRIEFING**

Despite its widespread use, the effectiveness of debriefing is far from clear, with relatively few controlled studies existing that examine its impact or effectiveness (Raphael, Meldrum & McFarlane, 1996). Studies that have been conducted demonstrate mixed findings with debriefing being found to have either no influence or produce negative outcomes, whilst others indicate debriefing is effective. Concerns have also been raised in relation to its effectiveness in preventing future problems, the effectiveness of group versus individual debriefs and the effectiveness of mandatory versus voluntary participation (Litz et al., 2001).

Research in this area is difficult to conduct and severe methodological problems limit the interpretability of most published studies (Deahl, 2000). The field is polarised, with strongly held views both for and against the process. Robinson (2000) and Mitchell & Everly (1995), for example, state that there is “overwhelming circumstantial and anecdotal evidence for the effectiveness of properly administered debriefings” (p. 6). They go further to state that the dramatic expansion of CISD through organizations around the world is testament to the fact that it meets important needs and demonstrates there are a number of features that do make it a preferred method (Litz et al., 2001). On the other hand, a systematic review of the available empirical literature on the effectiveness of one session psychological debriefing, conducted under the auspice of the Cochrane Foundation (Rose & Bisson, 1998), found no evidence to support the use of debriefing. They concluded that brief preventative interventions for trauma are inappropriate and that the current practice of routine psychological debriefing should cease.

In fact, the research evidence is inconclusive. Anecdotal reports and uncontrolled designs have often indicated some benefits from psychological debriefing (see Everly et al., 2000 for a review). Reports on civilian populations in non-controlled designs, indicate that rescue workers who underwent debriefing were able to assimilate their experiences more effectively (Raphael et al., 1983), and emergency services personnel who received CISD demonstrated less job turnover and fewer mental health problems (Mitchell & Bray, 1990). Further studies by Stallard & Law (1993), and Chemtob, Tomas, Law & Cremniter, (1997) reported reduced levels of symptomatology at follow-up, whilst other reports in Litz et al., (2001) found no significant reduction in symptoms over time.

Controlled studies on civilian populations tend to show minimal to no improvement and suggest that the value of debriefings may not be directly linked to its content but in the opportunity it allows for review (Solomon et al., 2000). Negative
implications of debriefing were further noted amongst a follow-up on individuals involved in the Newcastle earthquake (Bisson, Jenkins, Alexander & Bannister, 1997). Jenkins (1996), however found a positive effect of debriefing amongst emergency workers involved in a mass shooting with symptoms of anxiety and depression decreasing, however due to design faults, the results remain relatively inconclusive.

A point worthy of note is the consistent finding that participants in debriefings usually report high satisfaction ratings with the experience, even if it did not reduce symptoms in the long term (Robinson, 2000). Hytten and Hasle (1989) investigated the effects of debriefing on fire fighters, reporting that 66 percent found the intervention worthwhile and helpful, despite the fact that there was no significant decline in their anxiety levels. Whilst overall impressions suggest that debriefing is beneficial in making some individuals feel better at the time, it is interesting to note that this perception of benefit in the short term, as well as satisfaction with the opportunity to debrief, does not necessarily correlate with mental health outcomes (Solomon et al., 2000). It does however, provide an indication that the organization is concerned about the psychological well-being of its employees.

A major difficulty in the efficacy debate relates to the lack of consensus about the goals of debriefing. Some proponents argue that psychological debriefing is not designed specifically to prevent long term psychopathology, but to “mitigate the impact of a critical incident and to assist the personnel in recovering as quickly as possible from the stress associated with the event” (Mitchell & Everly, 1995, p. 8), with the aim to stabilise the current situation, minimise any detrimental effects, normalise experiences and restore an individual’s level of functioning. Debriefing is not intended to be a form of psychotherapy, thus it is unrealistic to expect such an intervention to reduce all symptoms associated with a stress reaction and Mitchell himself acknowledges that not everyone will benefit from a CISD, and that there will be times when individuals require more help than a debriefing alone can provide (Mitchell & Everly, 1995). On the other hand, it could be argued that if an intervention fails to improve longer-term adjustment following traumatic exposure its value is questionable.

Efficacy of Early Intervention in Military Populations

Military personnel, by the very nature of their employment, are at high risk of exposure to potentially traumatic events. Solomon et al., (2000), reports that between twenty and thirty percent of combatants will experience some level of psychological breakdown. It is therefore incumbent upon defence forces to address the potential for psychological sequelae as an occupational health and safety issue. The notion of psychological debriefing in military environments is not new. Soldiers have been “debriefed” by commanders after involvement in battle for centuries and it was thought that through the sharing of stories, soldiers would become better prepared for future encounters and morale would be enhanced (Litz et al., 2001).

During World War II, Brigadier Marshall from the United States Army developed a group debriefing process that became known as Historical Group Debriefing (HGD). Whilst its primary purpose was to gather historical data, it was also found to have
positive effects on those participating and assisted in maintaining cohesion and morale (Shalev, 2000).

Studies examining the effectiveness of early intervention in military populations are limited with those that are available reporting mixed results, following the same trend as studies on civilian populations (Solomon et al., 2000). For example, Ford, Shaw, Sennhauser, Greaves, Thacker, Chandler, Schwartz, McClain, (1993), investigated the use of debriefs in soldiers involved in Operation Desert Storm and found that debriefing was effective in resolving psychosocial symptoms in soldiers on their return home, with a further improvement noted at a twelve-month follow-up. However, due to the large range of uncontrolled variables, it is not possible to determine whether debriefing was the sole cause for these improvements.

Less encouraging findings were reported by Searle & Bisson (1992, cited in Deahl, 2000), who found that 66 percent of soldiers who were debriefed within five days of being exposed to a traumatic situation met criteria for a PTSD diagnosis five-weeks later. Whilst results from this study need to be interpreted cautiously due to the small sample size involved, the findings do support regular screening of soldiers and follow up post deployment given the multiple and often sustained nature of trauma associated with warfare. Likewise, Deahl, Gillham, Thomas, searle & Srinivasan’s (1994) study of Gulf War veterans found approximately 50 percent to have symptoms of PTSD nine months later, with debriefing making no difference to the incidence of PTSD symptoms.

Controlled studies on military populations generally conclude that debriefed groups are more likely to exhibit symptoms of depression and anxiety, however they are unable to determine whether the debriefing process is in itself damaging (Deahl, Srinivasan, Jones, Thomas, Neblett & Jolly, 2000). However the question arises over whether it is a process that is suitable for everyone.

While debriefing is a primary prevention strategy (administered “non-selectively” to the affected population), other secondary prevention strategies have been developed to assist those who have already started to develop acute (combat) stress reactions. During World War I, Salmon, a military psychiatrist, developed the principles of Proximity, Immediacy, Expectancy and Simplicity (PIES) to treat soldiers who were overcome by acute stress reactions such as “shell shock” and “battle fatigue” (Neria & Solomon, 1997; Shalev, 2000; Solomon et al., 2000). The PIES philosophy was to treat the soldier close to the battlefront (proximity), as soon as possible after the onset of symptoms (immediacy), with the expectation that they would return to duty within 24 –72 hours (expectancy). The intervention is basic, with a focus on the provision of rest, food, and opportunities for self-care and recuperation (simplicity).

Whilst empirical evidence on the effectiveness of this intervention is limited and generally only available retrospectively, it is a principle that is still adopted by some defence forces across the world. Preliminary research with the Israeli armed forces suggests its use in treating CSR may increase the rate of return to the line of combat, with up to 60 percent of soldiers returning to the front line after receiving this form of
intervention (Solomon & Benbenishty, 1986). However, its effect on reducing long-term morbidity remains unclear (see Solomon & Benbenishty, 1986; Neria & Solomon, 1999; Shalev, 2000; Solomon et al., 2000).

Several other strategies have been adopted by other defence organizations. For example, the US Army routinely provide a range of “psychological first aid” interventions to members exposed to potentially traumatic events. These may include informal discussions to help process the event or pastoral intervention (Litz et al., 2001). Other defence forces have adopted similar interventions and promote rest, as well as the use of psychoeducation in individual or group settings. Although empirical support is lacking at this stage, pharmacological interventions, notably anti-adrenergic agents such as clonidine (Friedman, Charney & Southwick, 1993) or beta-blockers such as propranolol (Pitman, Sanders, Zusman, Healy, Cheema, Lasko, Cahill & Orr, 2002) are sometimes used to assist in managing acute symptoms, particularly those of high arousal.

**PSYCHOLOGICAL MORBIDITY**

Whilst some argue that the risk of exposure to potentially traumatic events is high, particularly for emergency service personnel or members of defence forces (Kessler, Sonnega, Bromet, Hughes & Nelson, 1995), the prevalence of PTSD occurring in individuals exposed to such events appears to be, in most cases, relatively low. Results of the National Comorbidity Study (Kessler et al., 1995) indicated that eight percent of individuals exposed to traumatic events developed PTSD as a result of their experience. Such results were supported by Breslau, Kessler, Chilcoat, Schultz, Davis, & Andreski, (1998), who reported a rate of nine percent.

Most individuals who are exposed to a potentially traumatic situation will experience some level of symptomatology after the event, however symptoms usually subside within three months. Those individuals who continue to experience difficulties after this time are likely to be at risk of later development of PTSD or other psychological disturbance. Thus it would appear that whilst many people will experience some distress after exposure to an event and may require assistance in coping and adjusting to the situation, in most cases, the majority will not require ongoing assistance (McFarlane, 1988; Raphael, Wilson, Meldrum, & McFarlane, 1996).

Individuals who do develop PTSD or other psychological morbidity should take into consideration that a number of factors are thought to contribute to the prevalence of PTSD, including the nature of the traumatic event, intensity and duration of exposure, previous psychiatric history and coping style. Other studies suggest that chronic disorders arising post event are caused by a range of factors and that debriefing cannot be used as an accountable factor alone (Litz et al., 2001). Furthermore, given that between twenty and thirty percent of the civilian population are thought to suffer from one or more psychiatric disorders (Raphael et al., 1996), it is likely that in some cases, early intervention such as debriefing may not be effective enough on its own to prevent delayed onset of symptomatology, particularly when follow-up is often not provided (Raphael et al., 1996).
EARLY INTERVENTION CONSENSUS FORUM

Given the paucity of strong empirical data to support early interventions following trauma, a strong reliance must be placed on expert consensus. In this context, the “Mass Violence and Early Intervention Consensus Workshop” which was held in Washington from October 30th – November 1st, 2001, was the first of its kind to assemble a group of experts from around the world with the goal of developing a consensus document regarding world’s best practice in early intervention following trauma and disaster. The consensus forum identified several “Key Operating Principles” which indicated that key components of early intervention include preparation and planning, education and training for providers, and service provision for those affected. The importance of early assessment was addressed and recommended that these should focus on a hierarchy of needs (such as safety, security and food before addressing health services available or communicating with family). Key aspects of early mental health intervention were also identified and recommended the inclusion of assessment, monitoring recovery, provision of information and training, fostering resilience, coping, and recovery (especially facilitating the use of naturally occurring support networks) and providing further treatment for those in need.

ALTERNATIVES TO PSYCHOLOGICAL DEBRIEFING

Several components of a standard debriefing model are routinely adopted as part of a broader approach to disaster and trauma management. In view of the multiple uses to which the term “debriefing” is applied, and the many components that may comprise a psychological debrief, it is recommended that the use of more specific descriptive terms be adopted for each aspect. It then becomes easier to develop specific intervention models that provide a “best fit” for any given situation. Based upon evidence available on the efficacy of psychological debriefing, and that suggested by the Early Intervention Consensus Forum, a number of key components have been identified as being important factors in early intervention. These factors have been influential in the development on the ADF’s revised approach to responding to potentially traumatic events, the Critical Incident Mental Health Support (CMS) model and include:

Education and Information

It is routine practice in acute trauma management and disaster recovery (Creamer, 1996; Lystad, 1988), as well as in the treatment of more established disorders such as PTSD and ASD (Bryant, Harvey, Dang, Sackville & Basten, 1998; Foa & Rothbaum, 1998) to provide some simple education about common psychological responses to trauma. Such an intervention serves to allay fears of abnormal reactions and reduce emotional valence associated with acute symptoms. It is proposed that a better understanding of normal reactions to trauma serves to reduce avoidance behaviour and social withdrawal, prompt the use of adaptive coping strategies and facilitate recovery. Although there is an argument for providing this information prior to the occurrence of a traumatic event as part of routine training, there is often a place for providing such information again following the trauma. Equally, it is important that this information is presented in the context of an expectation of recovery in order to reduce the risk of “priming” the development of symptoms.
Similarly, it is common practice to provide information, where possible, about the incident and its consequences. Understanding and memories of the event are frequently confused, fragmented and distorted, making it difficult to integrate and assimilate the experience. Figley (1985) has noted that recovery from trauma is characterised by attempts to answer fundamental questions of “what happened and why did it happen?” The provision of accurate information facilitates understanding of the experience, helping to reduce negative appraisals of the event and feelings of vulnerability, and increasing perceptions of control. However, care should be taken not to overwhelm the person and information should be titrated according to need.

Social Support

Increased availability and use of naturally occurring social support networks is routinely associated with improved subsequent adjustment and evidence exists regarding the benefits of social support in enhancing recovery following exposure to a traumatic event (Creamer, Burgess, Buckingham & Pattison, 1993; Gist & Lubin, 1999; Soloman & Bebenishty, 1986; Ursano, Grieger & McCarroll, 1996). Thus it would be recommended that a primary task in providing psychological first aid following a critical event would be to facilitate the development and use of naturally occurring support networks.

Management of Acute Symptoms

There is increasing research suggesting that two acute symptom presentations – high arousal (Bryant, Harvey, Guthrie, & Moulds, 2000) and peritraumatic dissociation (Friedman, Brandes, Peri, & Shalev, 1999; Fullerton, Ursano, Epstein, Crowley, Vance, Kao & Baum, 2000) are associated with poorer adjustment.

The management of high arousal has been a central component of treatment for acute traumatic stress conditions (Bryant et al., 1998) and PTSD (Foa & Rothbaum, 1998) and arousal reduction strategies have been commonly used in many disaster recovery settings (Creamer, 1996). A reasonable body of strong empirical data has been accumulated to support the use of anxiety management in traumatic stress conditions (Foa, Keane, & Friedman, 2000), thus, teaching simple arousal reduction strategies is a logical component of early intervention.

The management of acute dissociative states is slightly more problematic. It is assumed that dissociative symptoms develop as a protective strategy against intolerable stress and caution should be exercised in attempting interventions targeted at reducing dissociation. No research has yet attempted to determine the potential benefits of reducing acute dissociation on subsequent adjustment. Nevertheless, it is to be expected that, if other early intervention strategies are implemented effectively, dissociative states are likely to be reduced. At the very least, the presence of peritraumatic dissociation should be noted as a high risk factor, prompting the need for close monitoring of the individual.
Confronting the Traumatic Experience

There is no doubt that some kind of prolonged therapeutic exposure to the traumatic memories is a treatment of choice for chronic traumatic stress conditions such as PTSD (Foa et al., 2000). What is less clear is the benefits of encouraging individuals to go through this process in the immediate aftermath of trauma and in the absence of effective screening for risk factors, acute symptom profile, and suitability. This is considered one of the potential dangers of a debriefing approach.

Given the evidence to support this approach (i.e., reviewing the traumatic experience in both formal and informal settings), as well as the apparent desire of many trauma survivors to talk about their experiences in the immediate aftermath, it may be best to consider this intervention as a possibility for some survivors after some incidents. A clinical decision should be made regarding the suitability of this approach in each setting, however, there should be no expectation that a detailed review of the experience and associated emotional reactions would be a routine procedure in every case.

CBT Interventions for ASD and PTSD

As noted above, one of the primary goals of early intervention is to identify those in need of more intensive intervention. It is beyond the scope of this paper to provide a detailed review of specialist psychological and pharmacological interventions for ASD and PTSD, however research evidence exists to support the use of structured cognitive behavioural interventions in the treatment of both acute (Bryant et al., 1998; Foa, Hearst-Iked & Perry, 1995) and chronic (Foa et al., 2000) mental health conditions. It is of vital importance that psychological first aid protocols emphasise referral where appropriate for more intensive treatment and that those options are readily and easily accessible.

Timing of Interventions

CISD is generally conducted almost immediately following a traumatic event, however there is little evidence to suggest that this provides maximum benefit. Questions have arisen surrounding the timing of interventions and whilst in many instances, debriefing is suggested within 48-72 hours of an incident (Dyregrov, 1989), it is acknowledged by others that immediate debriefing is often not possible, nor desirable (Deahl, 2000). In many cases, individuals remain too distraught to process any type of intervention at this time. CBT interventions administered by Foa et al., (1995) and Bryant et al., (1998) found that sessions provided at least 10 days after the incident, continuing over four to five weeks produced significantly better outcomes than single session interventions. A review of the existing RCTs on the effectiveness of one session of psychological debriefing (Rose & Bisson, 1998), suggested that brief preventative interventions for trauma are indeed inappropriate.

Perhaps interventions should therefore be timed according to the specific needs of the affected population and delivered in a “stepped” approach. It is reasonable to assume that the primary goal in the first 24 hours should be one of containment – establishing safety, addressing basic needs, and minimising the escalation of overwhelming emotions. It is also an opportunity to conduct an assessment of the situation, screen for high-risk
individuals, and begin to develop an appropriate “best fit” intervention. Subsequent acute interventions (24 – 72 hours) should depend on the specific needs of the affected population and provide an opportunity for more in-depth assessment of individual needs and identifying those requiring more intensive intervention. The importance of linking this phase with effective follow-up cannot be over-emphasised.

**Screening and Assessment**

Given the uncertainties surrounding the benefits of early intervention, a strong emphasis on effective screening and assessment is essential. A considerable body of knowledge exists to inform judgements about who will, and will not, be at high risk of subsequent problems following traumatic exposure. These include a range of pre, peri and post-trauma factors. If early intervention does nothing else, it should assist in identifying individuals who should be targeted for more intensive follow-up.

Initial screening and assessment will be strengthened by the use of standardised assessment instruments. Such psychological tools assist in providing a conceptual structure that can help an individual (and clinician) understand traumatic responses and provide critical information for subsequent follow-up. Whilst it is generally not a factor considered to significantly impact on an individuals level of distress it is one that warrants further consideration and could be used in conjunction with other processes. While several good measures exist to assess acute symptom levels there are few available to assess broader risk profiles. Preliminary risk assessment scales are currently used by the British Royal Marines and a scale is currently under development at the Australian Centre for Posttraumatic Mental Health (ACPMH).

Effective initial screening and assessment is also important in determining the suitability of specific interventions. Preliminary evidence suggests that debriefing may not be effective for individuals with depressive tendencies, as they may be likely to ruminate on the event (Shalev, 2000; Solomon et al, 2000). Furthermore, individuals who rely on avoidant coping techniques, as opposed to healthier active coping strategies, may become overwhelmed if they are unable to use their defensive techniques. Others at risk include individuals who have experienced multiple trauma (due to the risk of reactivating former reactions), as well as individuals displaying symptoms of ASD or other signs of high distress, such as CSR (Solomon et al., 2000).

**THE ADF MODEL OF EARLY INTERVENTION**

Consistent with the general community, mental health has become an increasing priority for the ADF. As a result the Directorate of Mental Health, which is a multidisciplinary team with key stakeholders from all areas of mental health has been established within the Defence Health Service Branch (DHSB). The development of the DMH coincides with the launch of an ADF Mental Health Strategy (ADFMHS), which recognises that mental well-being plays a key role in an individual’s overall health and that mental ill health may have an effect on a member’s ability to function at an optimal level. The strategy seeks to reinforce the concept of ‘wellness’ in life for ADF members and their families.
Trauma management has been identified as a priority in the ADFMHS, largely due to its implications for ADF operational effectiveness. Exposure to potentially traumatic events is a likely part of military service and for some individuals the type of response provided could mean the difference between a positive outcome or long term debilitating mental health problems. As an organisation, the ADF has a responsibility to take the ‘least harm’ approach and avoid providing interventions that may cause harm. As such, their approach to trauma management has been revised, with an emphasis on developing a practical but research based approach.

Based upon evidence presented in this paper, recommendations for acute intervention following exposure to potentially traumatic events in the ADF were proposed. Following a workshop with key stakeholders in 2002, an agreement was reached that the ADF would utilise a tri-service approach to managing responses to critical incidents, and the framework for an ADF model of Critical Incident Mental Health Support (CMS) was developed. The resulting model is based on the principles of ‘World’s Best Practice’ and ADF combined practical experience. The model seeks to provide the highest quality training for those who provide first line responses, in order to minimise the impact of critical incidents or potentially traumatic events and maximise ADF capability. The development of the model was a collaborative project between the Australian Department of Defence and the ACPMH and provides systematic, multi-level, standardised training across the three services (i.e. Army, Navy and Airforce).

The ADF model of intervention is based on the principles of World’s Best Practice as previously outlined, specifically that most people will recover and that interventions should be provided based on the nature of the incident and reactions of persons involved. The new ADF model shares many principles that were underlying in the previous ‘CISM’ based model (i.e. the importance of social support, basic education, and basic self-management strategies). It also provides a detailed structure in which stepped levels of intervention are provided, based on the concept that events will be more or less potentially traumatising, thus the nature and length of intervention and follow-up should be determined by how traumatising an event is for those involved.

The ADF CMS model is divided into two parts: Background (covering the history of trauma management in the ADF, a review of current practice and implications for the future), and Intervention and Skills development (based on key principles of early intervention as outlined in Table 2). The intervention component includes four phases, which cover planning and immediate responses, psycho-education and screening, and two follow-up phases, as outlined in Appendix A.

Table 2: Key Goals of Early Intervention

<table>
<thead>
<tr>
<th>Key Goals of Early Intervention Following Exposure to Critical Events:</th>
</tr>
</thead>
<tbody>
<tr>
<td>?? Facilitate access to and use of social support networks</td>
</tr>
<tr>
<td>?? Ensure basic needs are met (safety, sustenance, rest, contact with others)</td>
</tr>
<tr>
<td>?? Provide information about follow-up arrangements and available resources</td>
</tr>
</tbody>
</table>
Create an environment that re-establishes a sense of control
Screen for high risk individuals (pre, peri, post-trauma factors)
Provide assertive follow up for high risk individuals (or groups)
Provide basic information about common responses (reassurance)
Facilitate access to basic information about the event (what, why)
Reduce arousal
Improve basic self care (reduce stimulants and drugs, promote rest, good diet, exercise)

The CMS model is designed to provide a number of functions. These include:
- A high level of flexibility to accommodate the vast array of incidents and circumstances that may occur within the ADF;
- Include follow-up and on going care;
- Include a process by which the provider assesses the needs, provides intervention, reassesses and refers to a qualified mental health practitioner when required;
- Include liaison with Commanding Officers;
- Maximises existing support mechanisms and promotes adaptive modes of coping;
- Screens for risk factors and chronic problems; and
- Include coordinated support to families through a multidisciplinary approach.

Whilst this new model for providing mental health support to potentially traumatic events has been developed based upon key components of early intervention and expert consensus, it is recognised that this field is still largely in its infancy. Whilst some of the implications for practice are definite with their value being well supported by research, others have been established as probable, while the evidence regarding a third group of interventions has been inconclusive in relation to their potential risks and benefits. The ADF has a responsibility to rely on current research and formulate policy and practice in the field accordingly, and is now in a position to be able to evaluate the effectiveness of this approach to early intervention and influence future outcomes.

CONCLUSION
Overall it should be considered that the short term success of any technique does not guarantee a long term effect, thus the use of debriefing should be considered in line with other appropriate measures. It is recognised that there may be risks associated with debriefing and other forms of early intervention and it remains uncertain whether debriefing actually reduces long term psychological morbidity. Individuals who receive debriefing should not be considered to be immune from developing psychological sequelae after the events, thus follow-up from debriefing processes is vital. The role and function of debriefing should also be made clear, not only to those providing it but also those receiving it. Whilst its short-term benefits are recognised and can assist in helping workers return to duty, it should not be stated that debriefing will prevent future psychological discomfort.

Although strong empirical evidence is lacking, there is a mounting body of clinical and research data upon which to draw conclusions and guidelines for intervention following...
Such evidence suggests that early intervention is more likely to be appropriate if it is provided within the context of a broader stress management program and on-going organisational support. Thus, a comprehensive approach to the management of traumatic reactions that includes screening, education and follow-up is advised. Furthermore, individuals involved with providing such intervention should be trained and protocols implemented and carefully followed. The above conclusions have influenced the ADF’s approach to trauma management and the development of the CMS model. In developing a model for the provision of mental health support following critical incidents, the ADF are now in a position to contribute to the further understanding and development of world’s best practice in this field.
REFERENCES


APPENDIX A: ADF CMS Model

POTENTIALLY TRAUMATIC EVENT

Phase One: Planning and Immediate Response

Phase Two: Option 1 - No Response Required at this Time

Phase Two: Option 2 - Initial Screening and Psycho-education

Phase Two: Option 3 - Group Interventions

Phase Three: First Follow-up

Phase Four: Second Follow-up

Event not assessed as traumatic

Event potentially traumatic

Response Requested

Event involving significant loss or intervention requested by commander

Significant Symptoms

Significant Symptoms

Minimal Symptoms

Minimal Symptoms

Referral to practitioner

Referral to practitioner

Event not assessed as traumatic

44th Annual Conference of the International Military Testing Association
Ottawa, Canada, 22-24 October 2002
NEW ZEALAND DEFENCE FORCE ON-GOING ATTITUDE SURVEY

Lieutenant Colonel A. McCone & Lieutenant P.K. Gibbons, RNZN
Headquarters, New Zealand Defence Force
Private Bag
Wellington
New Zealand
Alan.McCone@NZDF.mil.nz

INTRODUCTION

In medium to large organisations, having an accurate picture of employee satisfaction, workplace perceptions, and policy and practice effectiveness and efficiency can be very difficult. In military organisations, the ‘picture’ is often distorted because those who make policy and who administer practice are somewhat removed from the bulk of personnel but have perceptions based on when they were part of that bulk.

The Personnel Branch, Head Quarters New Zealand Defence Force (HQ NZDF), has a long-term measurement plan. This plan includes establishing internal and external measurement and monitoring systems. A number of internal Human Resource related measures have been developed and are reported via the Navigator tool in the NZDF Quarterly report. An external measurement of propensity to join has also been developed, and the results are also reported on the Navigator.

The On-going Attitude Survey (OAtS) was the next step in internal measurement. The aim was to measure the general attitudes of personnel to various facets of working in the military, resulting in a measurement of commitment to service and to provide feedback to policy makers.

Rationale

There are a number of reasons for wanting to measure internal attitudes and opinions. Knowing, as the result of systemic questioning, the problems faced by personnel and the general perceptions of service life allows policy makers to tailor policy and funding priorities. Before the introduction of the OAtS, excluding Exit surveys, only the Royal New Zealand Navy (RNZN) had an internal measurement system for serving or currently employed civilian personnel. Other measures, usually as a result of individual academic studies, were haphazard and the questioning was of dubious utility to NZDF level policy makers.

Retention

The NZDF has for the last five years suffered from shortages of personnel. Over twenty trade groups were at the point where further loss would result in the Service having difficulty in deploying to its full potential. Almost half of those that had left uniformed service over the last twelve months had less than five years service. For civilian employees, the figure was three and a half years. This combination of factors meant that even if the Services managed to recruit sufficient personnel, the lack of
growth and experience within the services would limit the total knowledge of the organisation.

For the NZDF to retain a good knowledge base, releases need to be controlled, which means that the people who leave are those who the Services would chose to leave. To control releases, the NZDF needs to be an employer of choice; meaning that personnel want to join and want to stay. To be an employer of choice, the NZDF needed to be able to state what factors of Service or which employment conditions of civilians are those that affect an individuals willingness to serve.

Only be systematically observing attitudes in relation to a well defined set of variables, and studying the impact of attitudes on service over a lengthy period can the NZDF start to properly target policy and practice development.

**Commitment to Service**

Organisational surveys often look at establishing causal links between conditions and attrition. In a military organisation, it is essential to establish the other side of the coin, what causes or encourages people to continue serving. The only way to establish this is through longitudinal research based on a common question set.

**Cohort**

Surveys delivered over time also allow for cohort tracking. There is real value for the military in discovering the changes that occur to attitudes over time. The best way to collect this information is to employ a standardised questionnaire administered to a particular graduating class of recruits at spaced intervals throughout their careers. In-depth knowledge of the changes to attitudes again provides information for the targeting of policy and practice. To ensure that the information gathered is not an artefact of repetitive sampling, a comparison sample also needs to be collected. Making the cohort study a sub-set of an on-going attitude study is an effective way of achieving this.

**Feedback**

Organisations need to know whether the introduction of policies and practices has had a desirable impact. In the situation the NZDF is in, with changing organisational, operational and political requirements, it is often difficult to delineate weather any event or announcement has caused any adverse or favourable change in attitudes. The only way to track changes and determine what has caused change to attitude is through on-going surveying, with a consistent set of questions.

**Development of Commanders**

It is also possible, with this type of on-going surveying, to provide the organisation with feedback on the perceptions within services and/or particular units.

**DESIGN**

Surveying requires careful design. In this case, there were a number of conditions to be satisfied during the design process. A paradox of running on-going surveys is the maximising of response rates through good publicity and feedback to organisation and participants, while minimising the impact on the individual and organisation by making it an everyday part of working life.
Pre-conditions

Conditions that were identified that had to be satisfied during design were:

- Maximum response rate.
- Minimum disruption to existing programmes.
- Low administrative cost.
- Administrative ease.
- Identification of trends.
- Identification of impact of events.
- Full coverage HR issues.
- Capacity for investigation of specific events.
- Capacity for Service specific information.
- Quick analysis and reporting turn-around.
- Repeat sampling procedure.

The survey consists of three parts. That is, NZDF-on-going, NZDF-situational, and Single Service specific.

**NZDF**

The main part of the survey is a set of questions that were designed to isolate opinions about attitudes to a number of organisational variables. These variables are tied to a model of organisational commitment developed over a number of years by NZDF personnel. These questions will remain in place for at least twenty-four months before any editing. Some questions will be aimed at perceptions of conditions of service, some at experience, and some at intentions. The responses to the perceptions and experiences questions will be used to establish relationships with intentions information. Service information on the individual will also be tracked in perpetuity to determine relationships between responses to all three categories and actual longevity of service.

**Situation Specific**

In each survey will be a space for additional questions on matters of the moment. If a particular policy is in development, questions about a range of options might be asked. Perceptions of practices surrounding the policy issues can be measured before and after policy introduction. If there is an indication of a particular external event that will impact on the NZDF, pre and post measurement can also be undertaken about attitudes likely to be affected. An example of a specific situation to be explored is reasons for the low numbers of civilian personnel adopting the civilian Superannuation Scheme.

**Single Service**

Single Services may have a need to run questions that are Service specific and do not relate to the other Services. For this purpose, three versions of the survey will be produced, the RNZN, Army and RNZAF versions each having an appended Service specific section. It is possible that this will negate the need for individual Service measurement, except where a ‘snapshot’ of large numbers of respondents at once are needed.

**Publicity**

A small publicity campaign will be launched to ensure that personnel have had the opportunity to read about the aims of the survey process and the particulars of
administration, response, and feedback. The campaign will be directed at the three Service through a variety of vehicles.

Topics Covered

The model developed for question construction allows the survey to provide information on attitudes and experiences in the areas of:

- Senior Leadership.
- Supervision.
- Performance Management.
- Career Development.
- Change.
- Health and Safety.
- Job Satisfaction.
- Commitment.
- Equity.
- Family and Quality of Life.
- Career Intentions.

Non-Invasive

The whole sampling, administrative and data collection methodology was designed to be as non-invasive as possible. Because the sampling process is tied with ATLAS, the NZDF Human Resource Information System, it is also possible to utilise the biographical data from the electronic records to obtain information about rank, age, gender, ethnic status, service, etc. Thus, all the respondents will not be required to complete demographic type data but will only be faced with a survey centred on their experiences and perceptions, none of which will be specific enough for there to be worries about identifying the respondent. Personnel are advised that there is a tie to NZDF data, and reassured that personal identifiers will not be released outside the study team.

Survey analysis will be reported in such a fashion that individuals cannot be identified. Where feedback is given for use in the development of commanders (should this occur), units will only be analysed when enough personnel have completed the form so that it is impossible for a commander to pinpoint individuals. There will be no compulsion for individuals to advise their commanders that they have taken part in the survey process, and there will not be any feedback from the study group to commanders to tell them who has taken part.

Feedback Mechanisms

A problem with surveys conducted within the NZDF in the past is that the results disappear into the bureaucracy and respondents do not know what has happened as a result of their efforts, or whether any action has taken place at all. At regular intervals, a summary of findings and associated actions will be published in Service newspapers. Respondents are informed of this when completing the survey.

Feedback will also be provided to Chief of Defence Force, single Service and Joint Head Quarter chiefs, and relevant Branch heads within HQ NZDF as part of an on-
going process. The nature of reporting and feedback will be in targeted reports. The content of the reports will depend on the question structure and the requirements of the recipients. Further on I will discuss examples of feedback provided to the single Services from the Baseline survey.

On-going Measurement

The aim and the key to the success of this survey is that measurement will be continuous. This is not a case of massive administrations of the same survey on an annual or bi-annual basis. A limited number of selected personnel will receive the survey each week. In this way, any changes in attitude due to internal or external events can be picked up as they occur. In any one week, no more than 60 personnel spread across three services and a range of bases will receive the survey. This reduces the number of personnel at any one time exposed to the survey, and should not contribute to any ‘survey overload’ problem. If a particular Service needs to conduct extensive surveying, members from that Service may not receive the NZDF survey for two weeks either side of the single Service administration.

Retention

The direct link to ATLAS will allow for tracking of personnel to measure the impact of attitudes on attrition. When an individual leaves, a tag will be placed on their survey response, which will enable retention to be matched with attitudes to determine causal effects.

A Willingness to Serve Index will be developed to measure attitudes towards commitment to service and retention data. It is a composite score drawn from data obtained from surveys of all personnel. It is hoped that the first Index score will be available by the second quarter 2003 report.

METHODOLOGY

The methodology chosen for this survey was aimed at gathering and providing solid information with the least possible inconvenience to individuals, the least possible disruption to single Service workload and information gathering programmes, and the least possible additional workload to DSHRR.

Sampling

Sampling is based on no individual, with the exception of cohort studies, receiving a questionnaire more frequently than once every thirty-six months. The base population will include all personnel (uniformed and civilian) within the NZDF who have completed induction training. Sampling will be done on a random basis, from the ATLAS database. Any individual chosen in the general sample would not be eligible for re-sampling within thirty-six months. On current numbers, around sixty personnel would receive the questionnaire each week, spread proportionately over the Services and HQ NZDF.

For cohort studies, one set of induction course graduates from one Service would be chosen each year to follow through their careers. Which Service, and which stream
will be decided by the Services. An alternate is to choose one Officer and one non-commissioned class from each Service in year one, and not reselect for the next three to five years. Cohort members would be surveyed on graduation, at the six, twelve, eighteen and twenty-four month points, and each eighteen months after this. Cohort members will still be eligible for the general sample, and if selected within four weeks either side of the cohort administration will be administered the general survey with cohort specific information included.

Administration

The Baseline survey and future surveys had/have three options for administration. They are pencil and paper, intranet, and internet.

**Pencil and Paper**

The traditional methodology for administering surveys within the NZDF is via posted paper based questionnaires. For those personnel who do not have access to the Intranet (determined through usage of mailbox statistics) or who are deployed, this will remain the distribution method.

**Intranet**

Almost all civilians, and over three-quarters of the total NZDF population have regular access to an NZDF Intranet connection. It is possible to deliver and administer survey instruments via electronic means. After filling out the surveys on their computer, the electronic form is returned and processed automatically. Not only does this reduce the administrative burden, it also minimises transcription errors.

**Internet**

Using the same programming as available on the Intranet, it is possible to deliver and administer surveys via the Internet. This method will be particularly useful for those personnel on deployments or on board ship or elsewhere not connected to NZDF LAN line. As the survey need not be filled in on-line, only data transmission charges are invoked.

Compulsion

To get the best use from this survey process, the highest possible response rate is required. One method of gaining a high return rate was to make it compulsory for individuals to complete and return the form. There are arguments both for and against this approach. The arguments against tend to reflect the attitude that compulsion can create negative sentiment against the survey in general and that a response that is deliberately false, or has had no thought put into it, is more damaging than a lower response rate. The arguments for are based on the perception that almost all respondents in non-voluntary surveys respond relatively accurately, and that larger response rates enable more sophisticated analysis. The final decision was made to incorporate some words from the Chief of Defence Force ‘strongly encouraging’ personnel to take part. This approach will be continued in the future.

Individual Feedback

No individual feedback was given from the results of the Baseline survey and it is still being decided for future surveys. However, it is quite possible that, after twelve
months, most higher level commanders (sub-unit level and above) can be given feedback on how personnel under their command view the situation in their unit, without disclosing information about which individuals responded. Several larger companies use this type of survey to provide managerial feedback, the IBM spring survey being a good example, although the NZDF survey is not aimed at asking how individuals feel about their direct supervisor’s performance.

The utility would be to point out differing levels of satisfaction between units and to make commanders aware of them. It is not likely that, even if it was agreed to, this could be implemented within the first twelve months of administration as acceptance of the survey and the veracity of results would need to be established first.

THE BASELINE SURVEY

The Baseline Survey was sent out to two thirds of all NZDF personnel, military and civilian. This included personnel on operations and overseas postings. A total of 3261 personnel responded to the baseline questionnaire. That was just over a third of each Service and HQ NZDF.

The response rate was higher for the on-line surveys with a 70% response. The on line surveys were also quicker to be returned than the paper surveys.

Some interesting population details from the baseline survey are:

a. All but 164 respondents provided a Service number, which meant that we had a direct link to the HRIS and therefore access to very good biographical data to help analyse the responses.

b. In general, the lower ranks, from Recruits through to Corporal equivalents, were underrepresented. This is a trend we have noted in other surveying within the NZDF. 846 responded from this group.

c. A greater proportion of RNZAF personnel responded than did from the other Services.

d. Ethnicity and gender proportions were representational.

e. Slightly more civilians than proportional responded.

The analysis showed that overall, the data from the baseline survey is capable of providing a fairly accurate picture of attitudes across the NZDF in for that particular time.

Basic Results From Baseline Survey

The OAAtS was designed to give an overall picture of attitudes across the NZDF including Job Satisfaction. The survey contains several questions that aim to measure respondents Job Satisfaction. This information can also be linked through to the HRIS to track levels of satisfaction compared to who actually leaves, and what specific areas of job dissatisfaction is it that encourages people to leave, or satisfiers that encourage them to stay.

Job Satisfaction
Some interesting results from the Baseline survey regarding Job Satisfaction are:

a. Generally, only about 1 in 16 people across the NZDF claim to have poor satisfaction with employment in the NZDF.

b. There are some differences across military and civilian personnel, with 9.5% of military and 5.6% of civilians expressing poor levels of satisfaction.

c. Across the Services, around 8% of Navy and 9% of Army and Air claim Poor satisfaction. This compares with 3% of HQ NZDF. Table 1 shows percentage of poor satisfaction by service and military/civilian.

d. In terms of overall satisfaction, we can say that the majority of people are relatively satisfied.

Table 1. Percentage of Military and Civilian Personnel Claiming Poor Satisfaction

<table>
<thead>
<tr>
<th></th>
<th>MILITARY</th>
<th>CIVILIAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>HQ NZDF</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>RNZN</td>
<td>9</td>
<td>5.5</td>
</tr>
<tr>
<td>ARMY</td>
<td>10</td>
<td>5.9</td>
</tr>
<tr>
<td>RNZAF</td>
<td>9</td>
<td>7.4</td>
</tr>
</tbody>
</table>

Morale

The Baseline survey also asked respondents about changes in levels of their morale in the organisation over the last twelve months. Table 2 shows that, across all the NZDF, morale slipped slightly over the last year. However, there were significant Service differences, with Air showing a larger decrease and HQ a slight overall rise. These results are not particularly surprising given the restructuring changes in the RNZAF during the latter part of 2001. Responses to a question about personal morale show it is relatively strong right across the Services.

Table 2. Self-Report Changes in Morale in Organisation Over the Last 12 Months

<table>
<thead>
<tr>
<th></th>
<th>INCREASE</th>
<th>SAME</th>
<th>DECREASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>HQ NZDF</td>
<td>30</td>
<td>43</td>
<td>27</td>
</tr>
<tr>
<td>RNZN</td>
<td>26</td>
<td>39</td>
<td>35</td>
</tr>
<tr>
<td>ARMY</td>
<td>25</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>RNZAF</td>
<td>12</td>
<td>25</td>
<td>63</td>
</tr>
</tbody>
</table>
Satisfaction With Current Job

In the Baseline survey respondents were also asked about their current level of satisfaction with their current jobs. It appeared that although there is some poor satisfaction levels recorded the majority of respondents reported that they were satisfied.

When looking at satisfaction and attrition it appeared that:

a. In HQ NZDF, of those actively searching for other employment, 40% had Poor satisfaction with employment in the NZDF and 80% had poor satisfaction with their current job.
b. In the Navy, of those actively seeking employment, 76% had Poor satisfaction with employment in the NZDF and 72% with their current job.
c. For Army personnel, 70% of those actively looking for employment had Poor satisfaction with employment in the NZDF and 65% Poor satisfaction with their current job.
d. For Air personnel, 77% of those actively looking for external employment had Poor satisfaction with employment in the NZDF, and 69% Poor satisfaction with their current job.

Influencers

As stated above the OAtS asks a series of questions about attitudes on a range of areas. For each area there is a series of questions, and at the end of each series of questions, respondents were asked to pick those items that had the most influence on their satisfaction.

The top two influencers from each section were:

a. Senior Leadership
   - The quality of communication between management and staff.
   - The receptiveness of senior management to ideas from staff.
b. Supervision
   - The influence of the supervisor on the workplace working well together.
   - The amount of trust that can be placed in the supervisor.
c. Performance Management
   - The quality of the work of the workplace/section.
   - The quality of the feedback provided by the supervisor.
d. Career Development
   - Appropriate recognition for the amount of effort put in.
   - The quality of career management.
e. Change
   - The impact of recent Govt decisions on personal morale.
   - The impact of Govt decisions on willingness to stay.
f. Health and Safety
   - The amount of stress in NZDF related areas of life.
   - The importance with which Health and safety in the workplace is treated.

g. Job Satisfaction
   - The amount of pride one is able to hold in ones work.
   - The amount of enjoyment gotten from working with others in the workplace.

h. Commitment
   - The apparent future prospects of the NZDF.
   - Whether what the individual is doing adds to a belief that they are helping their country.

i. Equity
   - That people are judged on their ability to do the job, not on their gender.
   - That people are always treated with courtesy and respect in the workplace.

j. Family and Quality of Life
   - Flexibility to enable people to deal with family demands.
   - The impact that demands of the workplace have on family / home life.

k. Career Intentions
   - Probability of getting a good job outside.
   - Amount of perceived negative change from present circumstances.

When looking at the above results for all NZDF respondents there are some differences within groups. When looked at by gender, females are more interested in flexibility for family demands. Males are more likely to cite demands of work interfering with family as an influence. Females also view getting constructive feedback as more influential.

When considering results by service, RNZN personnel are more likely to cite senior management being receptive to ideas as an influence. HQ NZDF personnel are less likely to indicate career management as an influence on satisfaction, and far more likely to indicate getting appropriate recognition as an influence. Service personnel are more likely to see compatibility with others in the workplace as important than HQ NZDF (civilian) personnel. Army and Air personnel are more likely to see gender equality as influential on their satisfaction.

Self-report, such as survey data, on what influences satisfaction is useful, but not necessarily always accurate. As part of the initial statistical analysis question responses were used to discriminate between those who were satisfied and those who were not. The items in the lists below are ordered by discriminating power.

a. Items that had discriminating power across all sub-populations were:
   Perception of quality of life compared with other organisations.
Perception of utility to individual in staying.
Perception of whether the NZDF is a good organisation to work in.
Personal support for the organisation.

b. Additional discriminating items for Civilian Personnel only:
Degree of agreement with appraisal standards.
The level of stress experiencing.
Impact of changes of the last twelve months.
Level of satisfaction with the appraisal system.
Level of satisfaction with career management.

c. Additional discriminating items for Military Personnel only:
Degree of confidence in decisions by senior leaders.
Level of recognition for effort put in.
Degree of belief in the future prospects of the NZDF.
Degree of belief that senior leaders have the required skills.

FEEDBACK TO SINGLE SERVICES AND HQ NZDF

Feedback to single services was provided through an Intra Service Comparison Analysis. The analysis gives comparative views across groups of each service. It is not a full interpretative analysis. That is it does not draw definitive cause and effect conclusions. The analysis was offered as a way for Services to identify high priority areas for further work and / or investigation through additional processes.

The analysis is divided into 13 sections, each section representing an area of statistically related questions from the survey. There is an initial breakdown of each question within the subject area by Military and Civilian groups. Subsequent analysis occurs within these groups, across Gender and Location (where numbers permit) for Civilian employees, and across Gender, Rank, Trade and Location for uniformed personnel.

There were two main charts used to present the information to the services. The main analysis consisted of a series of charts that indicated the mean of responses across the group and sub-group for each question. Each question has four response options, from Strongly Agree (1) to Strongly Disagree (4). Where responses are either evenly spread or normally distributed across this range, the mean for the question would be on or near 2.5. A dotted line on each chart shows this point.

All questions are statements that are either positively or negatively weighted. If personnel were mostly satisfied then response means for positively weighted questions would be near to one (1), and the means for negatively weighted questions near to four (4). In each of the following charts, response means that merit discussion are circled and comments appear in the accompanying text box.
Where the numbers of responses from different groups were quite low the groups were excluded from the analysis because of the distorting effect this can have on interpretation.

Chi-square tests were conducted to test the hypothesis that the row and column variables are independent, with a significant value indicating there is a relationship between the variables.

The Services were also provided with charts that used all combined scores for all questions to give a distribution for each area. This was also conducted for all NZDF.
SUMMARY

The NZDF Survey is designed to fill a gap in NZDF measurement systems. The output can be used to provide basis and priority for policy and practice development, to identify personnel issues, to conduct longitudinal research in the antecedents for attrition, to identify differences between components of forces, and to identify the impact of external events on internal attitudes towards service.

The survey is designed to cover all three services, both uniformed and civilian personnel, with a consistent set of questions, as unobtrusively and non-invasively as possible. Sampling is automatic using ATLAS sources, and as far as possible administration, collation, analysis and reporting will be automated. Costs are relatively minimal and the whole process can be conducted within set resources once the process is embedded.

The NZDF views the OAtS as an essential instrument in our strategic HR toolbox. It will assist in developing willingness to serve measures as well as determining policy focus and effectiveness of policy implementation.
MEASURING ORGANIZATIONAL CLIMATE IN THE CANADIAN FORCES.

Major B.F. Johnston, Captain K.J. Brown, Captain A.K. Cole & Ms S. Aggarwal
Operational Effectiveness and Leadership Section,
Director of Human Resource Research and Evaluation.

“Human motivation and behaviour remain the keys to combat effectiveness”

(Kellet, 1982, p. 336)

Introduction

Within the Director of Human Resource, Research and Evaluation (DHRRE), the Operational Effectiveness and Leadership Section (OEL) is responsible for the measurement of organizational climate in the Canadian Forces (CF). This paper aims to discuss this research with reference to three major projects: the Human Dimensions of Operations (HDO), the Unit Morale Profile (UMP), and the Organizational Readiness Index (ORI). Finally, future approaches to unit climate research and suggested guidelines for using unit climate measures within the CF will be discussed.

Organizational climate is an important measure of human dimensions in any organization. For example, in a recent study of the Canadian Forces (CF), Chambers & Dowden (2000) surveyed 3003 personnel about their reasons for leaving. They found that 42% (n=2633) of CF members decided to leave due to dissatisfaction with job stress, travel, and leadership issues. They also reported that, among the reasons for leaving, 47% of respondents were dissatisfied with the organizational climate of the CF. Why is this so? How could an understanding of organizational climate have contributed to better management of personnel and avoided such dissatisfaction and loss of experience and skill?

To answer these questions a number of aspects of organizational climate need to be addressed: (a) what is organizational climate, (b) what are the benefits of measuring it and (c) how can outcomes be applied.

Definitions of Organizational Climate

At the heart of organizational climate research is the interaction between people and the organizations they work for. This interaction can be viewed from two perspectives: either person-centred or organization-centred. Focussing on the person, Lewin (1935) postulated that human behaviour is a function of the interplay of personality and the environment (i.e., \( B = f (P*E) \)). Others have focussed on the environment: for example, Schneider (1987) in his “individual attribute approach” theorized that environments are a function of persons behaving in them, (i.e., \( E = f (P*B) \)). That is, organizational climate is the product of interpersonal interactions in the work environment.
A global understanding necessarily balances both perspectives. An examination of the literature highlights: the pre-eminence and uniqueness of the person-organization interaction, the organizational values that can be deduced from the quality of this interaction, and the impact of organizational values on the behavior, motivation and intentions of its personnel.

Because organizational climate is best understood when the interaction of the individual and the organization is taken into account, by extension, it provides a conceptual link between analysis at the organizational level and the individual level (Litwin & Stringer, 1968; Payne & Mansfield, 1973) it enables an understanding of the impact of person–organization interactions on the nature of the work environment; and an understanding of how these, in turn, affect individual behavior and motivation (Litwin & Stringer, 1968; Sims & La Follette, 1975; Ashforth, 1985).

Unit climate measurement within the CF will now be discussed.

**Measuring Unit Climate in the Canadian Forces (CF)**

Since 1988, the Army has been the principal sponsor of DHRRE’s unit climate research within the CF and so will be the focus of this next section. The Army recognises that a principal component of operational readiness and effectiveness is “a solid human foundation laid on the requirements of military professionalism” (Canada’s Army, p91). The Army is interested in “soft” measures that provide hard data relating to the human element (e.g., “morale”) of operational readiness and unit effectiveness.

Measuring military effectiveness through unit climate is not an easy task. It is, however, one of the few means available to the Army to gauge the human factors impacting on readiness and unit effectiveness and to identify potential problems and develop suitable solutions. In effect, what gets measured, gets managed.

The following section provides a brief outline of the history of unit climate measurement with reference to the HDO and the UMP. Finally, the ORI project will be discussed as an example of organizational climate research with a military unit with predominantly civilian personnel.

Both the HDO and the UMP projects owe their origins to the research of Wild (1988) who developed a model that described the human factors predicting operational readiness. It included antecedent variables (e.g., aptitude, motivation and combat proficiency) and mediating variables (e.g., leadership). Building on this foundation, Hansen and Reeves (1989) developed the Human Dimension Combat Readiness Index (HDCRI-X). They identified several key factors in their
readiness model (e.g., morale, cohesion, leadership, and ideology) that were not only measurable but that reflected the doctrinal understanding of unit effectiveness.

By the end of 1995 the Army had confronted significant challenges as a member of the United Nation’s peacekeeping effort. Experiences in Somalia, Rwanda and Bosnia emphasised the differences between the relative clarity of NATO operations in Germany in a Cold War world and the ambiguity and enforced passivity that characterized peacekeeping operations. Some of the major differences impacted the human dimensions of operations: for example, command and leadership, ethical decision-making, and maintaining morale in “other people’s wars” (Capstick, 2000, p83). In response, the Army expressed interest in developing a tool for commanders to measure the human dimensions of operational readiness in their units. Based on Wild’s Readiness Model, and the developmental work of Reeves and Hansen, Murphy and Farley (2000) consolidated the Unit Climate Profile and extended it through the inclusion of measures of stress and coping. The final product became known as the Human Dimensions of Operations (HDO) project and currently comprises five subscales: (a) the Unit Climate Profile (UCP), (b) Stress in Military Service, (c) Signs, (d) Coping, and (e) Demographics. The HDO has been administered to every contingent deployed to Bosnia since 1997 and to contingents deployed to Eritrea, Haiti, Kosovo and Afghanistan also.

The Chief of Land Staff has mandated a two-year period to establish the HDO as a value-adding tool for commanders on operations. Specifically, the data collected in this period is to validate both the Operational Effectiveness Stress Model and the Stage Model of Operations. Research conducted to date has found support for both models and will be described briefly in the following sections.

Operational Effectiveness Stress Model

As hypothesized by Wild (1988), Izzo et al. (2000) found negative correlations between climate dimensions (e.g., morale/cohesion and confidence in leadership) and physiological signs of stress. Recently, Farley (2002) confirmed the negative relationship between negative coping strategies and stress and strain. Further, he reported that social support (via task and social
cohesion) moderated this relationship. Also, positive coping strategies had indirect, and moderating, effects on strain via their relations to morale/cohesion and confidence in leadership.

Stage Model of Operations

Schmidtchen (1999) developed the Stage Model of Operations to provide a framework for understanding how personal and work intensity changes over the course of a deployment. Further, Schmidtchen proposed that these changes impacted the attitudes and behaviours of soldiers in theatre. Riley (2001) found support for the positive relationships between intensity, morale, cohesion, leadership behaviour, and confidence in leadership. In addition, the hypothesized intensity curve of the model was found for established deployments but not for new, immature deployments indicating the further study of the Stage Model of Operational Deployment is warranted (Riley, 2001).

Future Directions

Previous research has indicated that a discord frequently exists between senior leaders' perceptions of soldiers' attitudes (e.g., morale) and the soldiers' actual attitudes (e.g., Stouffer et al, 1949; Eyres, 1998; Farley, 2002). Following a recent review of the HDO, the UCP Calibration Scale was included for surveys administered to Warrant Officers and commissioned Officers. The scale is intended to assess leaders' understanding of their soldiers' attitudes and opinions about operational stress issues. The data from the Calibration Scale will be used to reduce this discrepancy by providing feedback to leaders that will help them to calibrate their ability to accurately judge the attitudes of their soldiers.

The Unit Morale Profile

While proven to be an effective tool, the HDO is quite specific in its target population (combat arms and combat service support) and environment (war-like conditions). It was considered inappropriate for units in the garrison environment. In response to an Army request in 2001, OEL developed the Unit Morale Profile to measure unit climate in garrison.

The UMP examines the relationship between the determinants and the indicators of morale (e.g., role stressors, motivation, job satisfaction, psychological distress, leadership effectiveness, and cohesion) by applying several sub-scales to gather attitudinal and perception-
based data at the individual, sub-unit and unit levels. The UMP aims to provide commanders with a global measure of the morale of their unit and enable direct remedial intervention.

To date, the UMP has been administered to six units: both combat units and combat support units in garrison. A French language version has also been administered, however, the data is yet to be analysed. OEL aims to administer the UMP to three more units over the coming months and so be in a position, with a suitably large database, to review the instrument and the effectiveness of individual items and scales early in 2003.

**Organizational Readiness Index**

OEL is currently developing a unit climate measure for use in a civilian-military integrated Maritime environment, the ORI. Due to an ageing workforce and a dwindling pool of suitably skilled prospective employees, the Canadian Navy’s Fleet Management Facility—Cape Scott (FMFCS) anticipates a high percentage of turnover of personnel during the next decade. The challenge confronting FMFCS is to determine how “operationally ready” its workforce is to meet the predicted future challenges; as it prepares to confront the challenges ahead, FMFCS has identified the requirement for a measure of alignment between the organizational climate and its strategic vision and has approached OEL to develop the measure. This endeavour represents a shift in focus for OEL, a shift from studying a principally military organization to one that is principally civilian: Cape Scott’s workforce numbers approximately 1100 military and civilian members, mainly in technical and trades personnel, with the a civilian to military ration of approximately 4:1.

DHRRE conducted a focus group at Cape Scott and condensed the 11 performance indicators contained in the Unit’s strategic vision into five strategic themes: Communication; Personal and Professional Development; Leadership; Team-Based Culture; and Quality of Work Life. Following an extensive search of the literature, published scales were identified that had been shown to measure the components contained in the strategic themes. These sub-scales have been aggregated to form the Organizational Readiness Index. Administration of the instrument at Cape Scott is scheduled for December 2002.

**Recommended Guidelines**
Organizations interested in following the DHRRE approach should note that the measurement of organizational climate in the CF conforms to several guidelines:

a. the enduring aspects of an organization are identified and addressed;

b. measurement aims to capture a global “snapshot” of the organization independent of structural influences built into the CF (e.g., individual positions in the chain-of-command);

c. because OEL’s instruments measure individual perceptions, analysis takes into account the influence of factors external to the unit (e.g., the impact of the CF’s policy on frequency of postings on a Unit’s morale);

d. operating environments (operations or garrison) differ and require measures appropriate to each environment;

e. intrinsic factors, such as ethos and traditions, differ between elements of the CF (land, sea, air or civilian) and limit the validity of inter-elemental comparisons; and

f. surveying personnel raises expectations that where problems are identified some remedial action will occur; agreement from the sponsor organization that these will be somehow addressed is important.

Conclusion

An understanding of organizational climate and its relationship to maintaining an effective capability is continuing to evolve in the CF. As awareness of these important human dimension tools grows among commanders and supervisors so does the demand for their administration. The challenges of the next few years revolve around developing these tools into a readily applied and easily interpreted commander-friendly process for measuring the softer aspects of performance and climate. The future challenges are to:

a. develop an organizational climate “toolbox” that can be readily applied to the unique characteristics of individual units;

b. make climate measurement readily interpretable and understood by commanders as a value-adding exercise;

c. integrate the climate measurement process so that it can be applied seamlessly to the same unit whether “on operations” or “in garrison”; and
d. exploit technology to reduce the turn-around time from data collection to report presentation.

References


Canada’s Army: *We Stand on Guard for Thee*, DND: DGPA Creative Services 97CS-2161


Chambers, N., Dowden, C. (2000). To Go or not to Go? A Content Analysis of Open-ended Questions regarding why Members are (Thinking of) Leaving the Canadian Forces. A qualitative Data Analysis of the CFAIQ and the 1999 Fall Insert Questionnaire


INTRODUCTION

Is the U.S. Army prepared to meet the personnel needs of the Objective Force? Do the junior soldiers have the knowledges, skills, and attributes (KSAs) required for success as leaders in the transformed U.S. Army? How can the U.S. Army manage noncommissioned officers (NCOs) to ensure high quality NCO leadership in this era of change? The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) initiated the NCO21 Project in FY99 to address these issues by examining the junior NCO promotion system to make recommendations to the Army. This paper provides an overview of the NCO21 research including the background initiatives, research methodology, and study measures (Knapp et al., 2002; Knapp, McCloy, & Heffner, 2002). The accompanying papers (Hess, Entin, & Miller, 2002; Kilcullen, 2002; Sager, Putka, & Knapp, 2002; White, 2002) provide further details on the measure development and the concurrent validation effort.

The NCO21 research focuses on mid-career promotions, to sergeant (E5) and to staff sergeant (E6), in order to achieve the greatest span of influence on the NCO corps. Under the current system, promotions are based on common soldier knowledge, general skills, and training with limited emphasis on leadership potential or MOS-specific knowledge. The proposed future-oriented promotion system adapts the current common soldier model to include additional measures and better predict performance in current and future environments.

The research approach was a systematic evaluation of current and future factors (e.g., environment, personnel characteristics, technology) suspected to influence 21st-century job performance. After an initial exploration of the issues related to NCO performance in the future (Rumsey, Busciglio, & Simsarian, 1997), the NCO21 research was conducted in three phases. In Phase I, a methodology for examining and validating job performance in the future was developed. In Phase II, a future oriented job analysis was conducted. Included in this job analysis was a projection of the future conditions in which soldiers would work, identification of the performance requirements and...
associated KSAs, and prioritization of the KSAs. Finally in Phase III, pilot testing and field validation of the predictor measures and criteria were conducted. The predictor measures were adopted, adapted, or developed for this research, as were the criterion measures.

**Identification of Future Job Requirements**

For most civilian or military occupations, the first step to redesigning the promotion system is a traditional job analysis. This approach involves interviewing employees to identify the tasks they perform and the KSAs required to perform these tasks. Because we were interested in the KSAs soldiers need for future missions, we could not use traditional job analysis techniques. In lieu of the traditional approach, we conducted a future oriented job analysis by (a) reviewing future-oriented documents, (b) interviewing military planners and futures experts, and (c) analyzing existing jobs believed to be similar to future jobs. Over 400 written sources including official military documents and contractor reports were reviewed for pertinent information. Interviews were conducted with more than 300 subject matter experts (SMEs) and soldiers in future-like jobs (e.g., digital force soldiers, signal soldiers, military police, and special operation forces). This approach identified the KSAs and performance components expected to be important in Objective Force jobs.

Once the potential KSAs and performance components were identified, two expert panels assembled to complete the analysis. The first panel consisted of senior NCOs and officers from different MOS who had in-depth knowledge about future military conditions and jobs. These experts reviewed information about future expectations, revised the list of performance components and KSAs, and ordered the KSAs based on expected importance to future job success. A second panel of personnel specialists also ordered the KSAs, and the results from the two panels were combined. The most important KSAs, listed in Table 1, were earmarked for possible assessment (Ford et al., 2000).

The data collection conducted as part of the future oriented job analysis also allowed for the extraction of general themes that will impact all Army personnel in the future (Campbell & Knapp, 2002). These themes are detailed below.

- **Transition to digital operations and an ever-increasing pace of adaptation of technologies** – The change to digital/computer based technologies and updating of these technologies will occur frequently and quickly, requiring training and adaptation.
- **Diverse missions and frequent deployments** – Soldier will be expected to deploy more frequently, deploy more quickly, and deal with a variety of missions including peacekeeping, humanitarian aid, and nation building.
- **Diversity of forces and mission-specific organizations** – The variety of fielded equipment, techniques, and training will range from traditional equipment to the very cutting edge, with reduced standardization across the force, or even parallel units.
Decentralized operations – Technology will allow for greater dispersion and missions relying on smaller, more autonomous units.

Changes in training structure, requirements, and delivery – Training will require more self-assessment and self-management, with an emphasis on advanced training delivery techniques.

Changes in youth population and recruiting needs – Need for high quality youth in the Army will increase at the same time that more opportunities will become available to these high quality youth, creating even greater recruiting competition.

Table 1
Measurement Methods by KSA

<table>
<thead>
<tr>
<th>KSA</th>
<th>PFF21</th>
<th>ExACT</th>
<th>SJT (X)</th>
<th>AIM</th>
<th>BIQ</th>
<th>SSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive Aptitude</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conscientiousness/Dependability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effort and Initiative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotional Stability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General Self-Management</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrity/Discipline</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge of Interrelatedness of Units</td>
<td>XX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leading Individuals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management of Battlefield Functions</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Need for Achievement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOS/Job Specific Knowledge/Skill</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral Communication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical Fitness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem Solving/Decision Making</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Directed Learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervising Subordinates</td>
<td>XX</td>
<td></td>
<td>XX</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training Others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing Skill</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>XX</td>
</tr>
</tbody>
</table>

Note. PFF21 = Personnel File Form 21. ExAct = Experiences and Activities Record. SJT (X) = Situational Judgment Test and Situational Judgment Test-X. AIM = Assessment of Individual Motivation. BIQ = Biographical Information Questionnaire. SSI = Semi-Structured Interview. ASVAB = Armed Services Vocational Aptitude Battery. Table adapted from Knapp et al., 2002 (Table 8-1, 124-125).

The primary goal of this project was to identify improvements for the junior (semi-centralized) NCO promotion system. Using the data gathered from the future-oriented job analysis, we had the information needed to compare the current NCO promotion system to a future oriented system. Using the current system, the soldier’s
commander and the battalion promotion board make recommendations for considering soldiers for promotion. In addition to the recommendations by the commander and the promotion board, the soldier’s score is determined by (a) awards, decorations, and achievements; (b) military education; (c) civilian education; and (d) military training. A future-oriented promotion system designed to accommodate the current common soldier model yet include additional measures to better predict performance in current and future environments is ideal because it allows for integration of the existing measures without significantly altering the promotion system which has been so successful. Our proposed future oriented system is designed to tap those KSAs that are identified in the future oriented job analysis, but are not measured well in the current junior NCO promotion system.

**Identify and Develop Measurement Instruments/Methods**

We conducted a literature review to identify existing instruments that might be used to measure the critical KSAs (predictors) and performance requirements (criteria). Relevant sources included research studies, instrument development projects (particularly those done for the military), and test publishers. Considerations in the final selection of measurement methods/instruments included the following:

- Coverage of the highest priority KSAs and full coverage of the performance requirements.
- Anticipated reliability and validity in an operational context, as suggested by previous research and experience with similar measures.
- Reasonable development, validation, and administration costs.
- Suitability of the KSA measures for use in a large-scale promotion system.

**METHOD**

**Predictor Measures**

Seven measurement techniques were used to assess the current and future-oriented KSAs (see Table 1). The self-report Personnel File Form 21 (PFF21) measured key selection components of the current promotion system: (a) awards, decorations, and achievements; (b) military education; (c) civilian education; and (d) military training. In the past soldiers have been found to accurately report this information. The remaining measures assess future oriented KSAs.

The Experiences and Activities Record (ExAct) has 46 self-report items that measure how frequently junior soldiers engage in tasks such as training others, acting as supervisors, and working with computers. The items were written to tap five KSAs (see Table 1). Based on a factor analysis of the item responses, three scores were derived from the ExAct: Supervisory Experience, General Experience, and Computer Experience.

The Situational Judgment Test (SJT) measures eight KSAs including leadership and decision making (see Table 1) by presenting several brief scenarios and asking the
soldier to identify the best and worst possible actions from a list of four alternatives. Some of the 40 test items were adapted from other measures developed in other Army Research projects (e.g., Project A [Campbell & Knapp, 2001], Expanding the Concept of Quality in Personnel [ECQUIP; Peterson et al., 1997], Platoon Leader Questionnaire [PLQ; Hedlund et al., 1999]) and others were developed specifically for this research. The items were scored in relation to item responses provided by SMEs. (Additional SJT information is available in Heffner & Knapp [2002], Putka, Sager, & Waugh [2002] and Waugh, Sager, & Putka [2002].) A second situational judgment test (SJT-X) intended to measure the futuristic KSA “knowledge of interrelatedness of units” used three lengthier scenarios. These scenarios were developed explicitly for this research based on the themes identified as important for the future. Like the SJT, the items were scored in relation to item responses provided by SMEs.

The Assessment of Individual Motivation (AIM; White & Young, 1998) assesses six temperament qualities such as leadership, adaptability, and interpersonal characteristics by asking soldiers to describe themselves. The AIM is a 38-item forced-choice format assessment tool that is currently used in an operational test by the Army. Similarly, the Biographical Information Questionnaire (BIQ; Kilcullen, Mael, Goodwin, & Zazanis, 1996) assesses eight characteristics including leadership, conscientiousness, and interpersonal skills by asking soldiers to answer questions about their attitudes and past experiences. The BIQ is a 154-item multiple choice format measure developed from several operational measures. (Additional AIM and BIQ information is available in White [2002], and Kilcullen [2002], respectively.)

The Semi-Structured Interview, conducted with E4 and E5 soldiers, was used to evaluate an alternative to the traditional format for asking questions during the promotion board. Senior NCOs, E7 to E9, were trained how to ask questions, write questions, take notes, and score respondents answers. E4 and E5 soldiers responded to 14 hypothetical situation and past experience questions designed to assess seven KSAs (see Table 1). Two more KSAs, communication and military presence, were assessed from the soldier’s responses to all questions. Two MOS specific questions were included in the interview if an interviewer and the soldier were in the same MOS.

Together, these measures allowed us to represent the current promotion system through the PFF21 and a potential future oriented promotion system through the remaining measures. This approach allowed us to compare how much we can gain in job performance with measures of some or all of the future oriented KSAs to the current promotion system.

Criterion Measures

Two types of criterion measures were developed to validate the current (PFF21) and future oriented predictor measures. Supervisors provided performance ratings on two types of rating instruments. They rated current performance on 19 dimensions as well as overall performance and senior NCO potential ratings. Supervisors also rated their soldiers on how well they would be expected to perform in six scenarios characteristic of
the future Army. The second type of criterion was a computer simulation of a Humanitarian Aid Mission. Performance data was obtained from computer-generated scoring as well as observers’ ratings. (Additional criterion information is available in Sager, Putka, & Knapp [2002] and Hess, Entin, & Miller [2002], respectively.)

**DATA COLLECTION**

The measures were *field tested* on 513 E4, E5, and E6 soldiers at three Army installations. Job performance ratings data were collected from supervisors of the E5 and E6 field test participants. The field test was critical for evaluating and refining the instruments and also for trying out the data collection protocols and database management procedures.

We used a *concurrent validation* design in which we collected predictor data on E4, E5, and E6 soldiers and criterion data from E5 and E6 soldiers. The E4 predictor data were used as a basis for correcting validity estimates for range restriction. Data were collected at seven sites from approximately 1,900 soldiers and 988 supervisors representing 122 MOS. Table 2 shows the sample sizes by pay grade and location. The goal was to collect ratings from two supervisors per E5/E6 participant. Roughly 70% of the soldiers had at least one supervisor rater; only about 30% had two raters.

**CONCLUSION**

The three phase research approach for defining the methodology, conducting the future oriented job analysis, and conducting the field test and concurrent validation allowed this effort to take a very systematic look at junior NCO job performance in the present and in the future. The accompanying symposium papers will provide the specific analyses and results found using this methodology. From all of this information, we can conclude that it is possible to significantly increase the prediction of junior NCO performance by including measures of the critical KSAs defined in this research. In the context of a concurrent validation in a research setting, the measures of temperament, AIM and BIQ, and the situational judgment test made the most impact on the performance prediction. It is these measures that comprise the Leadership Assessment Tool (LAT). Plans are underway for additional research that will more closely mirror operational conditions and help us get a better understanding of the best ways to use the LAT. We also plan to develop alternate measures of the most promising instruments. Tools that may not be used for promotions, per se (e.g., the ExAct, the performance rating scales), may have important applications in the NCO training and development process.

*Table 2.*

Validation Sample Sizes
<table>
<thead>
<tr>
<th>Grade</th>
<th>Hood</th>
<th>Bragg</th>
<th>Lewis</th>
<th>Riley</th>
<th>Campbell</th>
<th>Carson</th>
<th>Stewart</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>E4</td>
<td>82</td>
<td>62</td>
<td>56</td>
<td>36</td>
<td>89</td>
<td>59</td>
<td>65</td>
<td>449</td>
</tr>
<tr>
<td>E5</td>
<td>177</td>
<td>162</td>
<td>69</td>
<td>143</td>
<td>126</td>
<td>102</td>
<td>106</td>
<td>885</td>
</tr>
<tr>
<td>E6</td>
<td>135</td>
<td>58</td>
<td>67</td>
<td>68</td>
<td>91</td>
<td>75</td>
<td>63</td>
<td>557</td>
</tr>
<tr>
<td>Total</td>
<td>394</td>
<td>282</td>
<td>192</td>
<td>247</td>
<td>306</td>
<td>236</td>
<td>234</td>
<td>1,891</td>
</tr>
</tbody>
</table>

REFERENCES


A QUASI-IPSATIVE TEMPERAMENT MEASURE
FOR ASSESSING FUTURE LEADERS

Leonard A. White, Ph.D.
Senior Research Psychologist
U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Ave., TAPC-ARI-RS, Alexandria, VA 22333
whitel@ari.army.mil

INTRODUCTION

This paper presents data on the criterion-related validity of a self-report temperament inventory, called the Assessment of Individual Motivation (AIM), against measures of current and future Army non-commissioned officer (NCO) performance. The Method section describes the (a) sample, (b) criterion measures used in the validation, and (c) the AIM. The Results and Discussion sections present the basic validation results and their implications for future research to support AIM’s potential operational use as an indicator of NCO leadership potential.

Background: Assessment of Background and Life Experiences

In the early 1980’s the Army initiated a long-term research program known as Project A to examine the criterion-related validity of current and new predictors of Army enlisted job performance. A follow-up Career Force project investigated the criterion validity of these predictor measures against a comprehensive set of measures of non-commissioned officer (NCO) performance (see Campbell & Knapp, 2001, for more details on these projects). As part of this research, a new instrument called The Assessment of Background and Life Experiences (ABLE) was developed to reliably measure six constructs from the temperament domain (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990). Constructs were selected for ABLE based on an extensive analysis of the criterion-related validity of temperament/personality constructs against measures of job performance criteria. A total of 237 studies conducted between 1960 and 1984 involving 339 independent samples were reviewed. In a complementary, concurrent study, experts estimated the likely relationships between these temperament constructs and 72 criteria of successful soldiering from analyses of Army enlisted occupations. Based on this review and the expert ratings the most promising constructs for predicting important Army job performance criteria were targeted for development on ABLE.

12 This paper is part of a symposium titled U.S. Army Non-Commissioned Officer of the Future: Building a Leadership Assessment Tool presented at the 2002 International Military Testing Association Conference in Ottawa, Canada (T.S. Heffner, Chair). All statements expressed in this paper are those of the authors and do not necessarily reflect the official opinions of the U.S. Army Research Institute or the Department of the Army.
Subsequent longitudinal and concurrent criterion-related validation research involving over 70,000 Army personnel showed that ABLE scales are significantly related to the job performance and attrition of entry-level soldiers and first-line supervisory non-commissioned officers (NCO). In this research, the uncorrected validities of ABLE scales ranged from .15 to .30 against multiple performance criteria including measures of leadership effectiveness, personal discipline (i.e., staying out of trouble), job effort, and maintaining physical fitness and military bearing. In other studies, ABLE scales were used to identify constructs important to the successful performance of Marine Security Guards who provide security services at U.S. diplomatic facilities worldwide, Air Force trainees, Army Special Forces Personnel, and Army civilian supervisors (see White, Young, & Rumsey, 2001, for a summary of these findings).

These temperament-performance relationships generated much interest in evaluating ABLE’s potential for use in operational screening. However, the Army ultimately decided to cancel a planned implementation of ABLE as a preenlistment attrition-screen. A major factor in this decision was concern about ABLE’s coachability and susceptibility to deliberate faking that might lead to score inflation and undermine ABLE’s predictive validity (White et al., 2001).

Assessment of Individual Motivation (AIM)

In response, the Army Research Institute initiated a research program to investigate alternative approaches for measuring these job-related temperament constructs from ABLE with less fakability. One such approach developed under this program was a new measure called the AIM. AIM is a multidimensional, self-report instrument that uses a forced-choice item format to measure the constructs from ABLE with the goal of achieving greater resistance to deliberate faking and comparable (or higher) criterion-related validity.

An early focus of the AIM research, following the Army’s plans for using ABLE, was to evaluate AIM’s fakability, coachability, and utility as a preenlistment attrition-risk screen. In research samples AIM was predictive of first-term attrition with incremental validity over educational attainment and AFQT that are the primary screening tools used by the Army. Correlations of AIM with examinees’ race and gender are uniformly low and suggest that using AIM for screening would not adversely impact females or minorities. Results from directed faking experiments also indicated that AIM’s forced-choice format is less fakable than ABLE’s normative scales or other more traditional self-report measures of personality (Young, Heggestad, Rumsey, & White, 2000).

This paper presents research on the criterion-related validity of AIM against measures of current and future job performance requirements of Army noncommissioned officers (NCOs) at pay grades E5 and E6. Within the set of AIM scales, it is hypothesized based on the Career Force findings, that Work Orientation and Leadership will have the highest correlations with measures of current and future NCO performance.
METHOD

Sample

The validation sample consisted of 1,442 E5 and E6 soldiers (i.e., \( n_{E5} = 885; n_{E6} = 557 \)). The AIM was administered in this research as part of a large predictor battery to soldiers stationed at seven locations in the continental United States. Performance ratings were collected for about two thirds of these soldiers. For more detailed information on the sample see Knapp (2002).

Performance Criteria

As described in the first presentation (Heffner & Knapp, 2002) in this symposium, two types of supervisor performance rating scales were developed for use as performance criteria in this research. The first was a set of 27 Current Performance Rating scales for measuring important aspects of soldiers’ current job performance. The second was a set of six Future Performance Rating Scales tapping raters’ assessment of soldiers’ likely performance futuristic scenarios. For this set of measures a future-oriented job analysis (Ford, R. Campbell, J. Campbell, Knapp, & Walker, 2000) was used to construct six scenarios describing conditions NCOs would be likely to face in the future. These scenarios covered topics like the increasing requirements for technical skill, self-direction, and self-management.

A single Current Performance Rating was calculated for each soldier as the mean supervisor rating across the 27 scales. Additionally, a single Future Performance Rating score was calculated for each soldier as the mean supervisory rating across the six Future Performance rating scales. The Current Performance interrater reliability estimates for E5 and E6 soldiers were .53 and .59, respectively. The Future Performance interrater reliability for E5 and E6 soldiers were .40 and .46, respectively. At the conclusion of the validation data collection, Current Performance rating scales had been completed for 1,001 soldiers and Future Performance Ratings for 1,012 soldiers.

Assessment of Individual Motivation (AIM)

The AIM administered in this research contained 38-items for measuring six psychological constructs from ABLE: Work Orientation, Adjustment, Agreeableness, Dependability, Leadership, and Physical Conditioning. Definitions for these constructs are shown in Table 1.
Table 1. AIM Construct Definitions

<table>
<thead>
<tr>
<th>Construct</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Work Orientation</strong></td>
<td>The tendency to strive for excellence in the completion of work related tasks. Persons high on this construct enjoy challenging activities, and set high standards for themselves. They consistently work very hard to meet these high standards.</td>
</tr>
<tr>
<td><strong>Adjustment</strong></td>
<td>The tendency to have a uniformly positive affect. Persons high on this construct maintain a positive outlook on life, are free of excessive fears and worries, and have a feeling of self-control. They maintain their positive affect and self-control even when faced with stressful circumstances.</td>
</tr>
<tr>
<td><strong>Agreeableness</strong></td>
<td>The tendency to interact with others in a pleasant manner. Persons high on this construct get along and work well with others. They show kindness, while avoiding arguments and negative emotional outbursts directed at others.</td>
</tr>
<tr>
<td><strong>Dependability</strong></td>
<td>The tendency to respect and obey rules, regulations, and authority figures. Persons low on this construct are more likely have difficulty following rules and get into trouble with the authority figures for violating rules or the law.</td>
</tr>
<tr>
<td><strong>Leadership</strong></td>
<td>The tendency to seek out and enjoy being in leadership positions. Persons high on this construct are confident of their abilities, speak up when they have something to contribute, and succeed in persuading others. They feel comfortable directing the activities of other people, and are looked up to when decisions have to be made.</td>
</tr>
<tr>
<td><strong>Physical Conditioning</strong></td>
<td>The tendency to seek out and participate in physically demanding activities. Persons high on this construct routinely participate in vigorous sports or exercise, and enjoy doing hard physical work.</td>
</tr>
</tbody>
</table>

Each AIM item consists of four statements (tetrad) that are descriptive of examinees’ past experiences and behaviors pertinent to the target construct. For each item, examinees are asked to identify which of the four statements is most and least descriptive of them. Thus, a valid item response yields four construct scores. A quasi ipsative scoring method is used that allows respondents’ total scores to vary depending upon the pattern of responses and thereby avoids some of the major psychometric problems associated with fully ipsative measures (Hicks, 1970).

Two goals guided the development of AIM (White & Young, 1998). The first was to reduce the impact of social desirability on item responding by attempting to balance the statements within a tetrad in terms of social desirability. Toward this end, items were revised (or dropped) if they were selected as “most like me” or “least like me” by more than 69% of examinees or were highly correlated with measures of socially desirable responding. The second goal was to make the content of AIM’s self-descriptive statements as behavioral as possible. Past research using rational biodata scales suggests that behavior-based items measure personality constructs with greater criterion-related validity, although internal-consistency reliabilities are usually somewhat lower as compared with more traditional self-report personality inventories (Kilcullen et al, 1995).
RESULTS

Table 2 presents the internal consistency reliabilities for the AIM scales for soldiers in pay grade E5 (n=434) and E6 (n = 537).

Table 2. Internal Reliability of AIM Scales by Pay Grade

<table>
<thead>
<tr>
<th>Scales</th>
<th>E5</th>
<th>E6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjustment</td>
<td>.69</td>
<td>.70</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>.74</td>
<td>.70</td>
</tr>
<tr>
<td>Work Orientation</td>
<td>.73</td>
<td>.69</td>
</tr>
<tr>
<td>Leadership</td>
<td>.65</td>
<td>.61</td>
</tr>
<tr>
<td>Dependability</td>
<td>.57</td>
<td>.55</td>
</tr>
<tr>
<td>Physical Conditioning</td>
<td>.64</td>
<td>.64</td>
</tr>
</tbody>
</table>

Table 3 shows raw and corrected criterion-related validity coefficients for E5 and E6 soldiers against supervisory ratings of current and future performance.

Table 3. Raw and Corrected Correlations between AIM and Criterion Scores by Pay Grade

<table>
<thead>
<tr>
<th>AIM Scale</th>
<th>Raw Observed</th>
<th>Raw Expected Future</th>
<th>Corrected Observed</th>
<th>Corrected Expected Future</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Performance</td>
<td>Performance Composite</td>
<td></td>
<td>Performance Composite</td>
</tr>
<tr>
<td></td>
<td>E5</td>
<td>E6</td>
<td>E5</td>
<td>E6</td>
</tr>
<tr>
<td>Dependability</td>
<td>.11</td>
<td>-.01</td>
<td>.12</td>
<td>.01</td>
</tr>
<tr>
<td>Adjustment</td>
<td>.06</td>
<td>.07</td>
<td>.05</td>
<td>.12</td>
</tr>
<tr>
<td>Work Orientation</td>
<td>.28a</td>
<td>.09</td>
<td>.28</td>
<td>.11</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>.01</td>
<td>-.01</td>
<td>-.01</td>
<td>.02</td>
</tr>
<tr>
<td>Physical Conditioning</td>
<td>.11a</td>
<td>.02</td>
<td>.10</td>
<td>.04</td>
</tr>
<tr>
<td>Leadership</td>
<td>.22a</td>
<td>.06</td>
<td>.26</td>
<td>.08</td>
</tr>
</tbody>
</table>

Notes. \( n_{E5} = 471-613; n_{E6} = 341-399 \). The correlations were corrected for criterion unreliability and range restriction on the predictor. Statistically significant correlations are bolded, \( p < .05 \) (one-tailed).

DISCUSSION

In this research several temperament scales from AIM were significantly correlated with supervisory ratings of current and future NCO performance. The highest correlations were found for Work Orientation and Leadership at the E5 level. The results also reveal some differences in AIM’s criterion-related validity for E5 versus E6 soldiers. Several proposed explanations for this unexpected pattern of validities by pay grade are currently being investigated.
Taken collectively, the results indicate that several scales from AIM have potential as additions to the U.S. Army promotion process to pay grades E5 and E6. Initial findings indicate that several have AIM scales have incremental validity over measures used by the Army in the promotion process (Sager, Putka, & Burnfield, 2002). Additional research is needed to establish AIM’s predictive and incremental validity in this context as a further step toward assessing its potential value as an indicator of leadership potential for entry-level NCO.

REFERENCES


Using U.S. Army Special Forces Biodata Measures to Predict 21st Century Non-Commissioned Officer Job Performance\(^{13}\)

Dr. Robert Kilcullen  
U.S. Army Research Institute for the Behavioral and Social Sciences  
5001 Eisenhower Avenue, Alexandria, VA 22333-5600  
kilcullen@ari.army.mil

INTRODUCTION  
The U.S. Army Research Institute’s NCO21 research program seeks to identify 21\(^{st}\) Century NCO performance requirements as well as the individual characteristics predictive of these performance dimensions. This research will lay the knowledge base for a promotion system geared towards the needs of the 21\(^{st}\) Century Army.

It is clear that NCOs in the 21\(^{st}\) Century U.S. Army will be faced with operating in progressively more complex and unstructured environments than ever before. Instead of confronting a global opponent whose capabilities and tactics are well known, they must now be prepared to face a variety of possible regional enemies, each with their own set of tactics and capabilities, under diverse weather and terrain conditions, on very short notice, anywhere in the world. This represents a quantum increase in the level of ambiguity and complexity that NCOs must be able to handle.

Technological improvements are also adding to the complexity of this task. Future military operations will require military leaders to precisely coordinate multiple ground, sea, air, and possibly space-based forces across the entire depth of the battlefield. Moreover, the rapid tempo of 21st Century warfare may require delegation of more decision-making authority to NCOs, and the lethality of new weapon systems increases the chance that the actions taken by these individuals will have a critical impact on the battle’s outcome.

One approach for developing valid predictors of future NCO performance is to investigate predictors of performance of soldiers who are currently operating in extremely complex, fast paced, unstructured environments. One such group of soldiers is the U.S. Army Special Forces soldiers.

The Special Forces play an important role in protecting American interests around the world. The Special Forces, commonly known as the Green Berets, consist of highly trained soldiers who perform a wide variety of missions, including counterterrorism, reconnaissance behind enemy lines, training/directing indigenous forces in guerrilla warfare and foreign internal defense, providing humanitarian aid to foreign countries, and executing assaults that seize/destroy targets or recover friendly personnel. As representatives of the United States, SF soldiers serve as diplomats, teachers, and peacekeepers as well as warriors.

---

\(^{13}\) The views, opinions, and/or findings contained in this paper are solely those of the author and should not be construed as an official Department of the Army or Department of Defense position, policy, or decision, unless so designated by other documentation.
Special Forces (SF) is designed around 12-man operational detachment teams (ODAs). These teams often operate for long periods of time under harsh conditions, isolated from other Army units. In order to perform their missions, these teams must be resourceful and self-sufficient. They must be flexible enough to overcome unanticipated obstacles and adjust quickly to rapidly changing contingencies without higher-order guidance. The performance of SF soldiers in these situations defines success for SF as a whole. While the specific missions carried out by SF soldiers will no doubt be different from those assigned to 21st Century NCOs, both sets of missions will require soldiers to perform successfully in complex, rapidly changing, unstructured environments.

Biodata scales used to predict various outcome indices in the SF community were included in the NCO21 test battery. This paper describes the criterion-related validity of these scales for predicting the job performance of Non-Commissioned Officers (NCOs).

METHOD
Subjects
A total of 1,835 NCOs consisting of 445 E4, 883 E5, and 553 E6 soldiers participated in this research. A wide variety of Military Occupational Specialties (MOS) were represented in the sample. Criterion information was collected only on E5 and E6 soldiers. After accounting for missing data the criterion sample consisted of 1,001 NCOs.

Measures
Rational biodata scales that showed promise for predicting job performance, delinquency propensity, and leadership potential in SF (Kilcullen, Mael, Goodwin, & Zazanis, 1999; Kilcullen, Chen, Goodwin, Wisecarver, & Sanders, in preparation) were administered to the subjects. Rational biodata scales measure temperament characteristics by asking questions about the test-taker’s past behavior and reactions to life events. Developing a rationally scored biodata instrument typically involves identifying motivational constructs (e.g., Achievement Orientation) likely to predict the criterion of interest and writing items that sample behaviors believed to be manifestations of these attributes. Item responses are scored based upon their presumed relationship to the construct, and item scores are summed to form scale scores having substantive meaning. Previous research on rational biodata instruments suggests that rational biodata scales can reliably and validly measure their intended constructs (Kilcullen, White, Mumford, & Mack, 1995).

Eight biodata scales were used in this research - Hostility to Authority, Manipulativeness, Social Maturity, Openness, Tolerance of Ambiguity, Interpersonal Skills, Emergent Leadership, and Social Perceptiveness. Two biodata scales that are strong predictors of SF mission performance – Achievement Orientation and Fitness Motivation – were omitted from this research since similar constructs are measured by the Assessment of Individual Motivation (AIM) test, which was also a part of the NCO21 test battery. The validity of the AIM temperament constructs is discussed in White (2002). Three of the biodata scales, Hostility to Authority, Manipulativeness, and Social Maturity are from the
Assessment of Right Conduct (ARC), a test designed to predict delinquent outcomes (Kilcullen, White, Sanders, & Hazlett, in press) and not job performance per se.

**Criteria**

Immediate supervisors rated the participants on one MOS-specific and 18 MOS-common dimensions of NCO job performance. Four outcome measures were derived – a single Overall Effectiveness rating, an Expected Future Performance rating, a Senior NCO Potential rating, and an Observed Performance composite derived from the 18 MOS-common rating dimensions. The development of these criteria is discussed in detail in Sager, Putka, & Knapp (2002).

Since the four outcome indices were highly interrelated, with correlations ranging from .75 to .85, the Observed Performance composite was chosen as the criterion in this research. The correlations of this composite with Expected Future Performance, Overall Effectiveness, and Senior NCO Potential were $r = .82$, $r = .85$, and $r = .81$ (all $p < .001$), respectively.

**RESULTS**

Descriptive statistics, internal consistency reliability coefficients (alpha), and intercorrelations for the biodata scales are presented in Table 1. Reliability coefficients were over .65 for all but the Interpersonal Skills and Tolerance for Ambiguity scales. The biodata scales targeted for predicting delinquency were highly correlated with each other (median $r = .61$), a finding that is consistent with previous research (Kilcullen, White, Sanders, & Hazlett, in press). Most other intercorrelations were more modest (median $r = .25$), although Interpersonal Skills was strongly correlated with the three delinquency predictor scales and Emergent Leadership was highly correlated with Social Perceptiveness.

Zero-order correlations were computed to assess the criterion-related validity of each biodata scale (see Table 2). Since the overall pattern of validities was reasonably consistent between E5 and E6 soldiers, the groups were combined to into a single sample to obtain a more stable indication of the scales’ predictive power. Except for Openness, each biodata scale significantly predicted the criterion (median $r = .11$, $p < .01$), although validities were relatively modest. Emergent Leadership was the best predictor of performance ($r = .18$, $p < .01$).

A multiple regression analysis was performed to assess the combined predictive power of the biodata scales. The subjects were randomly assigned on a two-thirds to one-third basis to the key-construction or the cross-validation sample. A backwards-elimination stepwise regression analysis was conducted on the key-construction sample using the biodata scales to predict the Observed Performance composite. Three scales uniquely predicted the criterion (see Table 3), Emergent Leadership, Interpersonal Skills, and Openness - the only scale that did not have a significant zero-order correlation with the criterion. A composite predictor variable
was calculated based on the regression analysis results in the key-construction sample and then applied to the cross-validation sample. A cross-validation multiple $R$ of .28 ($p < .01$) was obtained.
DISCUSSION

The results suggest that biodata scales used in the SF community show some promise for predicting 21st Century NCO performance. Still the magnitude of the prediction is modest. It is not too surprising that the biodata scales designed to predict delinquency (i.e., Hostility to Authority, Manipulativeness, Social Maturity) did not strongly predict the NCO job performance criterion. Somewhat more surprising is that the biodata scales that predict SF job performance showed more modest validities for predicting NCO performance.

Since SF soldiers are extensively prescreened with respect to mental aptitude, motivation, and physical fitness, it may be the case that the SF and NCO populations are too different for predictors of one group to work well for the other. It is also true that these groups work in different environments. The criterion in this research reflected NCO performance under current conditions – conditions that are less complex, unstructured, ambiguous, and volatile as SF conditions (and perhaps future NCO conditions). These factors may explain why the same predictors did not work equally well for SF and NCO soldiers.

In terms of assessing the importance of temperament for predicting NCO performance, it should be remembered that the two most powerful biodata predictors of SF performance, Achievement Orientation and Fitness Motivation, were not used in this research due to their construct overlap with the AIM. Therefore a better picture of how temperament relates to NCO performance would come from an analysis using the SF biodata and AIM scales in combination. Given that a cross-validated multiple $R$ of .28 was obtained using only the biodata scales, a more comprehensive temperament predictor battery likely would yield a cross-validated multiple $R$ in excess of .30. Therefore it seems safe to say that temperament plays an important role in determining NCO job performance.

REFERENCES


Table 1.

Descriptive Statistics, Reliability Estimates, and Intercorrelations for the Biodata Scales (n = 1,849)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>s</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Hostility to Authority</td>
<td>2.86</td>
<td>.58</td>
<td>(.72)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Manipulativeness</td>
<td>2.42</td>
<td>.55</td>
<td>.59</td>
<td>(.77)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Social Perceptiveness</td>
<td>3.53</td>
<td>.51</td>
<td>.09</td>
<td>-.12</td>
<td>(.82)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Social Maturity</td>
<td>3.34</td>
<td>.61</td>
<td>-.62</td>
<td>-.62</td>
<td>-.09</td>
<td>(.71)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Openness</td>
<td>3.37</td>
<td>.49</td>
<td>.14</td>
<td>.02*</td>
<td>.47</td>
<td>-.12</td>
<td>(.67)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Emergent Leadership</td>
<td>3.48</td>
<td>.49</td>
<td>.02*</td>
<td>-.18</td>
<td>.66</td>
<td>.05</td>
<td>.42</td>
<td>(.81)</td>
<td></td>
</tr>
<tr>
<td>7. Interpersonal Skills</td>
<td>3.15</td>
<td>.43</td>
<td>-.52</td>
<td>-.52</td>
<td>.18</td>
<td>.41</td>
<td>.07</td>
<td>.17</td>
<td>(.53)</td>
</tr>
<tr>
<td>8. Tolerance for Ambiguity</td>
<td>3.19</td>
<td>.40</td>
<td>-.32</td>
<td>-.38</td>
<td>.23</td>
<td>.22</td>
<td>.25</td>
<td>.29</td>
<td>(.45)</td>
</tr>
</tbody>
</table>

Note.
- Reliability estimates (coefficient $\alpha$) are on the diagonal.
- $^*$ $p$-value is not significant.

Table 2.

Bivariate Correlations of Predictors with NCO21 Observed Performance Composite Rating

<table>
<thead>
<tr>
<th>Predictors</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Hostility to Authority</td>
<td>-.12**</td>
</tr>
<tr>
<td>2. Manipulativeness</td>
<td>-.13**</td>
</tr>
<tr>
<td>3. Social Perceptiveness</td>
<td>.09**</td>
</tr>
<tr>
<td>4. Social Maturity</td>
<td>.11**</td>
</tr>
<tr>
<td>5. Openness</td>
<td>.00</td>
</tr>
<tr>
<td>6. Emergent Leadership</td>
<td>.18**</td>
</tr>
<tr>
<td>7. Interpersonal Skills</td>
<td>.14**</td>
</tr>
<tr>
<td>8. Tolerance for Ambiguity</td>
<td>.11**</td>
</tr>
</tbody>
</table>

Note.
* $p < .05$; ** $p < .01$

**Table 3.**

Cross-Validation Regression Analysis Predicting Observed Performance

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Std. Beta Weights$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Openness</td>
<td>-.09*</td>
</tr>
<tr>
<td>2. Emergent Leadership</td>
<td>.17**</td>
</tr>
<tr>
<td>3. Interpersonal Skills</td>
<td>.13**</td>
</tr>
</tbody>
</table>

Cross-validated R .28**$^b$

**Note.**

$^a$ key construction sample ($n = 1,292$)

$^b$ cross-validation sample ($n = 601$)

* $p < .05$; ** $p < .01$
PURPOSEFUL DEVELOPMENT OF FUTURE USAF LEADERS THROUGH OCCUPATIONAL ANALYSIS

Mr. Robert E. Boerstler, Civilian, Chief, Leadership Development Section
Mr. John L. Kammrath, Civilian, Chief, Occupational Analysis Flight
Air Force Occupational Measurement Squadron, Randolph AFB, TX 78150-4449
robert.boerstler@randolph.af.mil  john.kammrath@randolph.af.mil

ABSTRACT

This paper highlights the Air Force Occupational Measurement Squadron’s (AFOMS) involvement in the US Air Force Chief of Staff initiative for Developing Aerospace Leaders (DAL), now the Force Development (FD) Division within the Air Force Senior Leader Management Office. In order to provide current AF leadership with a strategy for purposeful development of future Air Force leaders, AFOMS was challenged to devise a means of maximizing “competencies” required for the general officers of the next generation.

Results thus far have provided initial guidance to the FD effort and set the stage for baselining the entire commissioned officer force structure through job analysis—something not previously done in the AF. Unlike the standard AFOMS analysis of the enlisted force, which focuses on work performed (tasks), this effort focuses on an officer’s knowledge and skills (competencies). Through personal on-site interviews with subject-matter-experts (SMEs), AFOMS has developed a complete list of competencies within the entire officer corps. Through the use of a web-based survey with various levels of branching, our initial job-typing would help define the best functional groupings of like specialties. Future work would focus on training and certification recommendations within these functional groupings.

The experience gained in performing this analysis, both in terms of methodology and web-based administration to a very large population, will provide a basis for process improvement for current and future customers.

INTRODUCTION

In December 1999, then Chief of Staff of the Air Force (CSAF), General Michael E. Ryan, was concerned about senior AF leaders attaining sufficient depth and breadth of experience to prepare themselves for leading the future force. He established the Developing Aerospace Leaders (DAL) Program Office to examine the current system of development for these leaders and to determine appropriate avenues for improvement. General Ryan’s concern: “Although our traditional, functionally managed career system was responsible for producing the world’s best Air Force, we had become an Air Force comprised of highly specialized competencies with too few airmen possessing cross-functional training or experience. This contrasted with our experience over the last 10 years, in which modern aerospace operations increasingly required effective, cross-functional coordination and smooth, horizontal integration within a larger mission
context. We realized we had not defined the institution’s larger leadership requirement, which would drive the development of all our officers under an overarching vision. Today, we have a great opportunity to provide our force with a more flexible, broad-based development path that directly addresses the institution’s future requirements—a path that will create airmen better prepared to serve and lead our integrated aerospace force well into the twenty-first century.”

The mission of the new Force Development (FD) Division (formerly the DAL Program Office) within the AF Senior Leader Management Office is to promote the deliberate and systematic development of future Air Force leaders through the implementation of innovative procedures and policies. To successfully create the best leaders possible, the Air Force must shift from the practice of managing career development from a predominantly functional or “stovepipe” system toward a more deliberate, broadened development process based directly upon institutional requirements.

Broadening a larger portion of the force at an earlier stage in their development will help ensure that individuals who reach senior-level positions are well-prepared to lead effectively. This preparation will include the proper balance of core competencies, career broadening assignments, professional education, training, mentoring, and deployments when applicable.

AFOMS is responsible for conducting occupational analyses for every enlisted career field within the Air Force and for selected officer utilization fields. AFOMS is an operational scientific organization that is often in contact the senior enlisted and officer career field managers through Utilization and Training Workshops (U&TWs). Occupational surveys generally provide information in terms of the percentage of members performing jobs or tasks, the relative percentage of time spent performing tasks, equipment used, task difficulty, training emphasis, testing importance (for enlisted specialties only), and the skills necessary to perform tasks. The structure of jobs within an occupation should serve as a guide for refining the Air Force military classification system or personnel utilization policies for that occupation.

With these capabilities, AFOMS was engaged by the FD office to assist in providing direction to attain the goals of the CSAF. Initially, AFOMS was invited to attend already established functional area Integrated Process Teams (IPTs). AFOMS provided guidance to better focus the IPT efforts and offered to collect and analyze data to support future classification and functional area restructuring.

**METHOD**

Competency lists initially developed by the FD IPTs were used as a foundation from which to develop a comprehensive list of competencies for all officer specialties. The competency inventory was developed and refined through interviews with subject-matter experts (SMEs) at operational units. Through this process, competencies were brought to an even level of specificity, which will allow for measurable outcomes and will enhance the analysis. Creating a competency inventory is a bit outside AFOMS’s historical paradigm as it includes the amalgamation of both knowledges and behavioral tasks.
AFOMS planned to survey all current officer specialties in pay grades O-1 through O-6 to identify which competencies are required for each specialty, and to what degree these competencies are required in the current Air Force structure. This structure analysis is designed to identify the current officer structure and make recommendations on possible restructure based on common competency requirements. Future recurring surveys would focus on specific specialties and would be designed to determine the training requirements for that specialty.

The survey instrument was set up to be web-based, interactive, and reentrant, with branching based on selected responses to facilitate ease of use for respondents and to provide analysis selection points. Requirements for expected outcomes of the data were defined, and strategies for implementation were planned. The analysis process was determined, and a timeline was established to meet these objectives.

The following analysis products were envisioned to provide empirical data to meet FD construct requirements:

**JOB ANALYSIS** – defines the types of jobs within the current officer structure, which can be identified by Major Command (MAJCOM) and base. Competencies within these jobs can then be analyzed for commonality with other specialties. This analysis can also help to determine the core competencies of the existing jobs, providing managers a useful tool to determine which competencies are suitable for certification within the specialty.

**AIR FORCE SPECIALTY ANALYSIS** – defines the competencies within the current specialties, with data reflecting competencies by grade or groups of personnel within a grade range. The data indicate differences between junior and senior level personnel within a specialty, which in turn assists in the analysis of progression within the specialty.

**MAJCOM ANALYSIS** – defines the competencies of specialties within a particular MAJCOM, which may or may not be compatible with similar jobs within other MAJCOMs. This analysis may be useful in determining competencies which may be considered for differences training between MAJCOMs.

**TIME IN SERVICE/COMMISSIONED SERVICE ANALYSIS** – defines the competencies expected at different points within a specialty based on total active federal military service or total active commissioned service. This analysis will be conducted in conjunction with the Air Force Specialty analysis to reflect the progression within a given specialty.

Specific background items have been submitted to AFOMS from the Systems Acquisition, Space Systems, Mobility Operations, and Information Operations communities for information which might be of benefit in the management these specialties. Although most needs can be addressed in the first survey to determine
structure of the entire officer corps, some items, such as requests for training requirements and methods of training, should be addressed in the recurring surveys for each functional area. Trying to include all items of interest to specific core specialties in this initial survey with such a large population will dilute the data by trying to determine too much from data scales not intended for that purpose. This survey must focus on occupational structure with future surveys addressing other specific concerns.

Due to fiscal constraints and a redirection of the FD effort, our plans to complete this structure analysis of the entire officer corps are currently suspended following our completion of the competency list development. Should future officer competency requirements emerge, the foundation that has been developed thus far will be instrumental in positioning AFOMS to quickly respond to future requirements.

**CONCLUSION**

The US Air Force is focused on developing a transformational leader who understands the full spectrum of air and space operations to include the development, support, employment, and sustainment of air and space power. AFOMS was enlisted to assist this purposeful development through occupational analysis and stands poised to support a comprehensive support base for officer development.

**REFERENCES**

Ryan, Michael E, General, Chief of Staff (Summer 2001), *Transformational Leaders – Aerospace Power Journal*, Volume XV, No. 2
The Automated Survey Generator (AUTOGEN)

Elizabeth J. Brady, Ph.D.
Senior Research Psychologist
U.S. Army Research Institute for the Behavioral and Social Sciences
brady@ari.army.mil

Introduction and Overview

One of the goals of the U.S. Army Research Institute for the Behavioral and Social Sciences’ (ARI) Occupational Analysis Office (OAO) is to continually improve the use of technology in studying Army jobs. In 1999, the OAO began working on improving the technology of their job analysis survey software. At the request of the U.S. Army Training and Doctrine Command (TRADOC), the OAO began creating an automated external training evaluation survey tool in August 2000 that could be incorporated into the current job analysis package. This software package became known as the Automated Survey Generator (AUTOGEN), with Version 1.0 distributed to the TRADOC schools and centers via the Internet in April 2002.

Background

When the ARI-OAO received the initial request for user-friendly standardized survey software that could be administered by TRADOC schools and centers, we contracted with the Human Resources Research Organization (HumRRO) to assist in gathering necessary information. Telephone interviews were conducted with TRADOC representatives at 26 schools and three centers to determine the bases of their current programs (i.e., what kind of surveys are they using, etc.). This information was taken into account along with the ARI-OAO’s work in selecting the best software to meet the Army’s needs.

Software Comparisons

Approximately 20 software packages were considered, with 10 ruled out for inability to support web-based surveys, inability to branch, or no statistical capabilities. Four software packages were seriously considered based on interview comments as well as first hand experience: Raosoft, Inc., Survey Tracker, Survey Pro, and Perseus. All of the software packages were compared using the following criteria:

- Survey administration by web, diskette, and Local Area Networks (LAN)
- Prevention against respondents taking the survey more than once
- Branching capability
- Number of survey templates

14 Presented at the annual meeting of the International Military Testing Association, October 2002. All statements expressed in this paper are those of the author and do not necessarily reflect the official opinions or policies of the U.S. Army Research Institute or the Department of the Army.
Number of questions in library
Ability to download files to temporary directory on PC to create an executable (.exe)
Variable, value labels are created automatically
Centering capability
Prename fields
Page name automatically changes
Allows the word “finish” to appear when document is complete
Color text
If-then logic supported
Statistical package included
Recoding capability
Control fonts/colors possible regardless of html, java, e-mail, & diskette
Force responses
Supports 1500-1800 questions
Database type
Compressed run time executable

ARI-OAO’s Recommendation

ARI recommended the use of Raosoft, Inc., and the Assistant Deputy Chief of Staff for Training approved this recommendation in February 2001. This recommendation was based on the following rationale:

Time investment
People investment
Proven track record
Current work supported
Development of an Automated Survey Generator and Handbook for on-going work
Purchase of “Unlimited Master Site License within TRADOC”

AUTOGEN Highlights

The AUTOGEN was designed specifically for the TRADOC schools and centers in developing, gathering, analyzing, and reporting information on job analysis and external training evaluation surveys. By using the AUTOGEN, surveys can be developed in a matter of hours versus the weeks or months it used to take. Once the survey has been developed, it can be sent out via floppy disks, E-mail, LAN, or the Internet.

No survey expertise or statistical skills are required to use the AUTOGEN. The system uses “templates” coupled with “fill in the blank” and “cut and paste” functions.
Additional questions may be added at the end of each survey. At the current time, up to 35 additional questions may be added. However, this number is expected to change in the next version based on TRADOC guidance.

The analysis portion of the AUTOGEN is “browser-based” with a standard package of reports available as well as the capability for the schools and centers to generate custom reports to meet unique requirements. Data may also be exported into Excel or SPSS for further analysis.

When requested, the ARI-OAO will host a school or center’s survey on its secure server. This server is available for use by the schools and centers free of charge.

A tutorial/help system is built into the system to provide assistance to the user at any time. Training time on the tutorial is estimated at a day or less.

The ARI-OAO is currently working with TRADOC representatives at the headquarters as well as the schools and centers to assist in their use of the AUTOGEN and to solicit their ideas for system improvements. The OAO will develop and maintain all future revisions to the software.

Version 1.0, which was released to the schools and centers in April 2002, focuses on the primary courses (e.g., Advanced Individual Training (AIT), Officer Basic Course (OBC), Warrant Officer Basic Course (WOBC), Basic Non-Commissioned Officer Course (BNCOC), Captains Career Course (CCC), Warrant Officer Advanced Course (WOAC), Advanced Non-Commissioned Officers Course (ANCOC)) for external training evaluation and job analysis surveys.

Version 2.0 will add functional course and common task survey capabilities. It is anticipated that this version will be released to the schools and centers by the end of December 2002.

A demonstration of the AUTOGEN will be provided at the conference. Please direct any questions you may have about the AUTOGEN to Dr. Elizabeth Brady, U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: TAPC-ARI-OA, 5001 Eisenhower Avenue, Alexandria, VA 22333-5600 or by commercial phone at (703) 617-0326 or DSN 767-0326.
Human Command And Control Functional Requirements for Future Systems

Mr. William R. Sanders and Dr. Carl W. Lickteig
U.S. Army Research Institute for the Behavioral and Social Sciences
Armored Forces Research Unit, Fort Knox, Kentucky 40121-5620
William.Sanders@knox.army.mil

ABSTRACT

The U.S. Army’s proposed Future Combat System of Systems (FCS) will be a networked force horizontally and vertically integrated from strategic to tactical level, to provide dominant situational understanding. Automated Command and Control (C2) capabilities in FCS units will allow tactical commanders, assisted by a small command group, to effectively lead a futuristic force composed of large numbers of manned and robotic elements. This paper describes research conducted by the U.S. Army Research Institute (ARI) to develop an approach for estimating human performance requirements associated with FCS C2 design concepts presented in experimental simulations. The Human Functional Analysis approach identifies and describes C2 functions associated with command group performance and assesses the duration and frequency of actions in support of those functions. Results of this analysis can be used to address issues of job design and task allocation, assess the effects of changes in automation support on workload, and serve as indicators of training and proficiency levels of the C2 element. Preliminary data were gathered in a series of U.S. Army warfighter-in-the-loop battle simulation experiments. Clusters of C2 tasks performed by the command group were identified from verbal protocol analysis (VPA), behavioral observations of Human-Computer Interaction (HCI), and focused surveys. Particular attention was paid to identifying the frequency and duration of verbal communications and soldier-system interactions associated with basic C2 functions (e.g., Plan, Move, See, Shoot). Discussion focuses on planned refinements to the Human Functional Analysis approach and its implementation in future FCS C2 research efforts.

Background

Research Goals. The research goals for this effort involve assessing the human performance required for command and control of future forces equipped with automated C2 systems. This required the development of measures of individual and collective performance to support task allocation, workload estimation, performance assessment, and training requirements. In support of these goals, ARI conducted a functional analysis of human performance across a series of future-oriented C2 experiments. This report describes some of the measurement approaches developed to identify human functional requirements for future C2 systems.

Human Functions Analysis. The analysis of human performance supporting C2 functions required the development of measures to assess command group verbal communications and HCI task accomplishment. For this paper the term “C2 Cell” will refer to a co-
located command group composed of a Commander, and three Battle Managers. The term “functions” is used to refer to groups of related actions that in combination serve to achieve a definite goal or purpose (Sanders et al., 2002). The FCS concept exploits technical advances in synchronized C² capabilities, intelligence and surveillance, maneuver, and fires. A review of U.S. Army documents addressing FCS C² functions (DARPA, 2001), and the U.S. Army Objective Force Operational and Organizational Plan for Maneuver Unit of Action (TRADOC, 2002) suggested that four basic C² functions could be identified (Plan, See, Move, Strike). These four functions provide a framework for the analysis of C² Cell verbal communications and HCI performance as follows:

- **Plan**: Develop, assess, and modify a plan including combat instruction sets provided to robotic elements in response to changing events.
- **See**: Control and interpret input from a heterogeneous set of advanced sensors to mentally construct an accurate picture of the battlefield in terms of METT-TC (mission, enemy, terrain, troops, time, civilians) factors.
- **Move**: Control the movement and activity of friendly manned and unmanned systems to maintain desired movement rates and formations.
- **Strike**: Distribute a variety of indirect and direct effects over a set of targets.

**Research Benefits.** A detailed assessment of C² functions and workload requirements can support many important decisions related to manpower, personnel, task allocation, materiel, and training requirements. For example, the HFA analysis provides useful estimates on the impact of C² prototype design changes on command group performance and effectiveness. Also, the behavior-based HCI measures provide an empirical basis for the development of automated C² performance assessment and feedback tools for training.

**MEASURES OF HUMAN C² FUNCTIONS**

The Human Functional Analysis approach adopted by ARI was designed to identify and describe the C² behaviors of the command group for a future fighting force. This effort developed detailed descriptions of critical command group functions, including operational definitions and behavioral anchors. Data supporting measures development were gathered in a series of U.S. Army warfighter-in-the-loop battle simulation experiments. A simulated C² Cell environment was developed for research purposes composed of a hardware and software system located in a command group C² vehicle. The simulated C² environment included workstations for four key command group players—Commander, Battle Space Manager, Information Manager, and Effects Manager—that allowed them to command and control a large number of robotic airborne and ground vehicle sensors, and other ground vehicles.

**Verbal Communications Analysis**

Verbal communications were analyzed to identify patterns and themes of communication. The analysis of command group verbal communications required the transcription of audio recordings of all spoken exchanges by members of the command
group with one another (Within Cell), with a Supporting Unit, and with higher headquarters (HQ). A taxonomy of communications was developed as a structural framework for a Verbal Communications Rating Scale as follows:

?? Source of communication: Within Cell, Cell-HQ, Cell-Support Unit, HQ-Support Unit

?? C² Function: Plan, See, Move, Strike

?? Type of communication: Share, Action, Direction, Ask, Process, Decide

?? Subject: METT-TC Factors (Mission, Enemy, Terrain, Troops, Time, Civilians)

A written transcript of the C² communications was developed from a video recording of battle runs. Teams of raters coded the transcripts to identify the Source, C² Function, Type, and Subject of the communications. Estimates of inter-rater agreement were calculated to ensure that the subjective ratings attained a high level of reliability.

The verbal communications measurement approach provided a format for summarizing a great deal of C² Cell performance information using a small set of C² performance dimensions. One assessment of C² Cell performance involved coding the verbal communications transcripts to identify the Subject, or information content, of the communications in terms of METT-TC Factors. Figure 1 presents the mean percentage of verbal communication time by METT-TC Factor across nine simulated battle runs lasting approximately one hour in length. In order by amount, the average percentage of total battle run time devoted to each METT-TC Factor was Troops = 29.8%, Enemy = 27.6%, Mission = 24.4%, Civilians = 3.2%, Time = 2.7% and Terrain = 1.3%. The Troops factor reflects the amount of verbal communication about friendly force assets, including weapon and sensor platforms. Notably, the relatively large percentage of communication time devoted to Troops may reflect the demand on humans for verbal coordination in the command and control robotic, semi-autonomous weapon and sensor systems.
The assessment of C² verbal communications performance was particularly valuable in identifying the lack of communication devoted to assessing the battlefield Terrain and Time requirements associated with the movement of forces. It was noted that the C² Cell had become very familiar with the battlefield terrain, and that this might be the reason that the C² Cell did not devote time to discussing this issue. Another indication of this is that the C² Cell members did not use the prototype terrain analysis and time estimation tools that had been developed for this purpose. A number of additional analyses of verbal communications were conducted. The percentage of run time devoted to communications between the C² Cell and the Supporting Unit, and between the C² Cell and HQ were also examined, to identify the subject matter and time devoted to communications across different levels of command.

**HCI Behavioral Observation**

Not all C² Cell tasks were reflected in verbal communications. The assessment of HCI performance provided an additional detailed account of the activities required to command and control forces in battle. The HCI analysis of player interactions with C² vehicle computerized systems required a review of video recordings of command group performance at each workstation in the C² Cell. A taxonomy of HCI tasks was first developed as a structural framework for an HCI C² Rating Scale. A written record of the HCI actions performed by the C² Cell Commander and the three Battle Managers was then developed from video recordings of one battle run. Teams of raters then coded the
transcripts to identify the type of HCI action performed, and the time duration. Estimates of inter-rater agreement were calculated to ensure that the ratings attained a high level of reliability. One goal in the HCI analysis was to promote the development of automated measures of command and control performance that would reduce the laborious process of video review, transcript generation, and task coding.

The primary measure of performance used for assessing HCI C\textsuperscript{2} performance was HCI task frequency. Task performance time and errors were less useful as performance criteria. Performance times for most HCI tasks were typically less than 5 seconds, did not differ across duty positions, and did not appear to be associated with any other indicator of success or failure. The HCI performance measures have been particularly valuable in identifying the effect of changes in automation support on workload, where workload was estimated by task frequency, for each member of the C\textsuperscript{2} Cell. Changes in C\textsuperscript{2} Cell automation support can occur when new features are added to system, and also when automated features are removed, or fail. While the introduction of automated features can potentially reduce workload, new requirements for set-up tasks or data entry associated with the automated features can actually increase workload levels. Figure 2 presents a breakout of the additional HCI workload resulting when target imagery analysis tasks previously performed by others were transferred to the C\textsuperscript{2} Cell. The frequency and percentage increase of tasks performed for the C\textsuperscript{2} Cell when the imagery analysis tasks were added was 23 tasks (11.7\%) for the Commander, 20 (5.1\%) for the Battlespace Manager, 68 (20.5\%) for the Information Manager, and one task (0.3\%) for the Effects Manager.

![Figure 2. Target Imagery analysis task contribution to C\textsuperscript{2} workload by duty position.](image)

The HCI performance measures can provide an emerging empirical basis for reallocation of tasks among command group players, and can identify priority areas for task automation. In general we would like to balance workload equally across the C\textsuperscript{2} Cell so that those members with available resources might assist in performing tasks for members who are heavily burdened. By breaking the hour-long battle into smaller intervals it was possible to identify the periods of high and low task demands for each C\textsuperscript{2}
Cell member. Figure 3 illustrates the frequency of HCI task performance for each of the C² Cell members during 9 ten-minute intervals of battle execution. Considering the HCI workload of the Commander as one example, he performed 219 HCI tasks during the 90-minute long battle simulation, which will be referred to as a “Run.” His greatest HCI workload occurred during the 50-60 minute time interval, during which he was heavily involved in sensor data display and target recognition tasks. His lowest HCI workload occurred during the 30-40 minute time interval in which his interactions were limited to a few sensor data monitoring tasks.

The HCI data provides an emerging empirical basis for reallocation of tasks among command group players, and can identify priority areas for task automation. As with the Verbal Communications assessment, the HCI assessment provides one perspective on task workload, which may not reflect all aspects of workload present. As example, while the Commander may have a low HCI taskload during some intervals of the battles, this might simply reflect the fact that he is required to perform high priority verbal communications, or decision making tasks at the same time.

Figure 3. Duty position HCI task load by battle time interval.

Interviews and Focused Surveys

Interviews and focused surveys were used in the present research to address issues such as training adequacy, workload, performance success, and C² feature effectiveness (see Table 1). Surveys asked participants to provide ratings and written comments to address specific topics. Surveys were administered according to a schedule, with training adequacy assessed at the conclusion of training, while assessments of system feature effectiveness and workload were administered in later runs after participants gained familiarity with the new features. Workload and Performance Effectiveness were
assessed at the conclusion of each Run to capture the impact of manipulations in battle Run complexity.

Table 1. Interview and Survey Measures Supporting C² Human Performance Assessment

<table>
<thead>
<tr>
<th>Measure Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-Training Survey</td>
<td>Adequacy of individual and collective skills training</td>
</tr>
<tr>
<td>In-Place AAR</td>
<td>Participants summary of “What went right and what went wrong”</td>
</tr>
<tr>
<td>After-Run Surveys</td>
<td>1. Run workload: Task Load Index (TLX)</td>
</tr>
<tr>
<td></td>
<td>2. New C² features effectiveness</td>
</tr>
<tr>
<td></td>
<td>3. C² Interface support of C² HCI tasks</td>
</tr>
<tr>
<td></td>
<td>4. New HCI features impact on workload</td>
</tr>
<tr>
<td></td>
<td>5. Interface support of C² Functions and METT-TC Factors</td>
</tr>
<tr>
<td></td>
<td>6. Write-in examples of teamwork skills</td>
</tr>
<tr>
<td></td>
<td>7. Write-in examples of C² decision making</td>
</tr>
<tr>
<td></td>
<td>8. Individual and collective skills proficiency</td>
</tr>
<tr>
<td></td>
<td>9. Human systems integration feedback</td>
</tr>
<tr>
<td>End-of-Experiment Interviews</td>
<td>1. Follow-up issues interview</td>
</tr>
<tr>
<td></td>
<td>2. Preliminary results review</td>
</tr>
</tbody>
</table>

The In-Place After Action Review (AAR) was used to get a quick summary of what went right, and what went wrong from each C² Cell member immediately after each battle run, while they were still in the battle simulator and being videotaped. These short summaries were quite valuable to video data reducers in identifying key events to look for in each run. The After-Run Surveys were conducted at the conclusion of each run and provided the researchers with an opportunity to gather self-report estimates of workload and performance success, and to address other issues such as equipment design changes. The End-of-Experiment Interview was typically conducted, providing an opportunity to gather additional C² Cell member input on issues raised during the course of experimental trials. The End-of-Experiment Interview also provides an opportunity to discuss preliminary findings with the C² Cell members for a “reality check”.

It is best to investigate issues from multiple perspectives, using multiple tools. In particular, the Focused Surveys provided an alternative method to address workload issues, which complimented the C² Verbal Communications and HCI workload criteria of task frequency and time duration. Figure 4 presents data on C² group member self-reports of performance success across three levels of battle scenario complexity, using the NASA Task Load Index (TLX) (NASA, 1998). Battle scenario complexity was manipulated by changing the level of friendly force automation support, and by changing the composition of the threat force. The TLX scale is a multi-dimensional rating scale that has been shown to be very sensitive to changes in operator workload levels in many different contexts. The After-Run Survey TLX Performance Success ratings generally fell at or above the scale value of 50, which marks the center point on the “Failure to Perfect Performance” 0 to 100 rating scale. Performance Success estimates appeared to differ across both player positions and scenario complexity. For all but the Effects Manager, there appeared to be a decrease in C² Cell member estimates of Performance Success at the Too High level of Complexity. This result matched the intent of the
experimental design for the Too High run which was to challenge the C² command group, identifying an upper performance ceiling for command and control performance.

![Graph showing performance success ratings by command group position and complexity.](image)

Figure 4. Average performance success ratings by command group position and Complexity (Medium, High, and Too High).

**FUTURE DIRECTIONS**

A number of refinements are planned for the Human Functional Analysis approach to support its implementation in future C² research efforts. These refinements should include the development of an automated data capture capability for HCI behaviors, and an effort to expand coverage of decision making and teamwork issues. A synthetic C² Cell task environment is currently being developed that will allow for focused and highly controlled experimentation to develop reliable empirical estimates of C² Cell team processes and performance outcomes.

The development of an automated data capture capability is essential for future research efforts, and could provide the basis for an automated performance assessment capability supporting training, evaluation, and C² system design. Manual video data reduction of command and control performance can only examine a fraction of the data potentially available from each FCS C² experiment, or any future FCS training, testing, and evaluation effort. Previous HCI video data reduction efforts have required approximately one day of analyst time for each hour of recorded performance. With eight separate video screens in the FCS C² Cell, data reduction has required approximately eight days to complete, identifying over 1,000 HCI actions. As a first step toward the goal of automated data capture, the Start and Stop actions associated with key C² HCI behaviors have been identified, which could allow software to be developed to pull this information from a data stream.

Future C² Human Function Analysis efforts should place a greater emphasis on developing decision making and teamwork process measures, as compared to battle outcome measures, which would include the activities, strategies, responses, and
behaviors employed in task accomplishment. Cannon-Bowers and Salas (1997) state that this distinction between process and outcome measures is essential in training development efforts, as outcome measures are usually not diagnostic, and do not indicate the underlying causes of performance necessary in providing constructive feedback. They suggest that process measures more directly describe the correct performance of interest, which is important because a team might achieve the desired outcome through a flawed process.

In support of ARI’s research on future C^2 Cell concept exploration and training, a multi-purpose in-house research laboratory is being developed at Fort Knox. The small in-house laboratory should serve as a site for the development of innovative C^2 concepts that can later be applied in larger test environments (Lickteig et al., 2002). The in-house laboratory will provide a small and highly controllable environment for exploring issues that have been identified, but cannot be addressed, in larger free-play battle test environments. The research environment will exploit the theme of situated performance, placing humans in a realistic digital representation of future C^2 system operational environments.

REFERENCES


NEW DIRECTIONS IN NAVY ASSESSMENT: DEVELOPING A MULTIMEDIA PROTOTYPE
Annette G. Baisden and AWC Bob Rymusaha
Naval Education and Training Professional Development and Technology Center
Pensacola, Florida, USA, 32505
annette.baisden@cnet.navy.mil

ABSTRACT
Technological advancements have transformed our militaries and have caused us to question the relevance of our assessment instruments to task performance. Emerging computer technologies afford opportunities to add to the range of assessment tasks we can deliver. The purpose of the present study was to prototype a multimedia test for potential use as a Navy enlisted advancement exam. Individual and group variation on different types of tests is well documented. This differential item functioning may be attributed to cultural context, cognition strategies and preferences. Multimedia tests provide a means to identify talents not identified by traditional assessments. Currently, multiple-choice questions assess verbal, quantitative, and analytical skills. The prototype was designed to include new interaction types that required test-takers to use spatial and auditory strategies for solving real-world problems. Handheld, tablet, and personal computers were evaluated for delivery options. Developing effective interactive multimedia assessment tools requires both psychometric evaluation and usability testing. Usability parameters included screen design and layout, feedback, system integrity and user satisfaction. Beta tests using authoring software for the Windows PC platform were administered to 86 enlisted aviation personnel. Debriefing sessions were held and survey data were collected to assess test-taker preferences and attitudes and to improve test design and effectiveness. User satisfaction was high with 82 percent of the test takers preferring the computerized format to traditional tests. The lessons learned in developing the prototype exposed unique issues and challenges in planning and designing tests that use innovative item types and laid the groundwork for future test development.

INTRODUCTION
New technologies and advances in cognitive and measurement science are changing large scale assessment. It is becoming more performance-based and dual-purposed, serving instructional purposes as well as high-stakes decision-making; occurring at a distance; and measuring skills newly valued by society (Bennett, 1998). The Navy produces excellent enlisted leaders, but its advancement processes are manpower intensive, unnecessarily slow, logistically difficult to administer and costly. Solutions to these problems require breakthrough process changes, including application of existing and future technology and emerging exam philosophies.

Transitioning to electronic testing has the potential to overcome many of the current assessment problems. Navy advancement exams are currently presented in paper-and-pencil format. Test questions are multiple choice and black and white figures or photos are used as needed. Questions test subject-matter knowledge and have little or no performance component. The exam process is significantly delayed by the return and receipt of answer sheets that can take up to six weeks. Electronic testing results are available immediately. Software databases and statistical reporting features provide the capability to streamline and tailor feedback to the Fleet. An electronic exam would enable a realistic presentation of subject matter, enable a
performance element, eliminate printing requirements, and reduce processing delays and security problems associated with handling of paper exams.

The Navy's approach to electronic testing involves three phases. Phase I includes the development, testing, and evaluation of an electronic exam prototype for the Aviation Warfare (AW) rating at NAS Jacksonville. Phase II extends the prototype development to other locations and ratings. Phase III requires future Navy-wide investment in computers/devices suitable for full-scale testing. This paper describes development efforts and lessons learned during Phase I.

PHASE I OBJECTIVES

The primary objectives during Phase I were development of a performance-based multimedia electronic exam, usability testing and formative evaluation. Performance-based assessment is the new challenge for assessment. The challenge is how to assess Sailors' understanding and competency through their performances on authentic and contextually meaningful tasks which require not just the recall of knowledge but also the use of knowledge in reasoning and problem solving. Using multimedia, problems can be created from real situations with appropriate levels of complexity. Navy advancement exams now test at the knowledge level of Bloom's taxonomy (Bloom, 1956). Multimedia improves the quality of assessment by providing practical application and testing at higher levels of analysis, synthesis and evaluation.

Usability is a key concept of human-computer interaction (HCI). It is concerned with making computer-based interactive multimedia systems user-friendly through a user-centered design process (Preece et al., 1994). Lee (1999) identified six dimensions of usability testing for developing effective interactive multimedia: learnability, performance effectiveness, flexibility, error tolerance and system integrity, and user satisfaction. From an HCI perspective, design must focus on the tool, the user, the tasks, and the environment (Schackel, 1991). Dillon and Zhu (1997) noted a lack of supporting evidence in the literature and stressed the importance of empirical methods to determine the most reliable and valid guidelines for interface design. Our goal was to use scientific methods to understand users and the nature of the task and ensure that human factors in interacting with computer were appropriately addressed in the design.

Formative evaluation is important because it provides quality control and focuses on improvement throughout the product development cycle rather than only at the end (Dick & King, 1994). Formative evaluation, also called prototype evaluation, could be considered as a theoretical base of usability testing (Lee, 1999). The goal of our formative evaluation was to obtain data to guide revision and to identify and solve usability problems before production.

PROTOTYPE DEVELOPMENT

The electronic exam team included AW subject matter experts, a personnel psychologist, instructional systems specialists, exam writers, computer programmers, and a systems specialist. The team started by reviewing electronic testing programs at other educational institutions, various software options and delivery methods. Perception for Windows by Questionmark was selected as an authoring tool primarily because of its widespread commercial use and its success in Navy training for end-of-course testing. The Navy was marketing the concept of a "Palm in the hand of every Sailor" so the initial evaluations of delivery options focused on handheld devices or personal data assistants (PDAs). Five brand names were tested using an eight-question multimedia prototype. Testing was unsuccessful due to hardware limitations such as poor resolution on small screens and limited storage capacity. Limited success was achieved with the Compaq iPAQ PDA using wireless transmission. Wireless transmission as a delivery method was not pursued further because of security deficiencies. Desired results were achieved with desktop computers in a LAN and Web-based Internet delivery. New design features,
technological advancements, and portability created an interest in evaluating tablet PCs as a delivery option. Hardware and software licenses were procured and a new prototype was developed for evaluation. The evaluation process and findings are described below.

**METHOD**

**Participants**

Two samples of enlisted personnel participated in the evaluation of the electronic exam prototype. Fifty five enlisted personnel from Commander Patrol and Reconnaissance Wing Eleven (CPRW-11) on NAS Jacksonville, Florida volunteered to participate in the first testing session of the study. The participants included all AW rates--from newly graduated AW "A" school AWAN's to AWCS. Thirty-one Sailors at the Afloat Training Group at Naval Station Mayport, Florida volunteered to participate in the second testing session. The participants were from the AW, Intelligence Specialist (IS), Operations Specialist (OS), Sonar Technician Surface (STG), Information Systems (IT), and Cryptologic (CT) communities. The participants did not receive any compensation for their efforts. Most of the participants were intrinsically motivated by the opportunity to take a practice test that might benefit their performance on future advancement exams.

**Materials and Equipment**

Data collection consisted of three parts: an interactive multimedia exam, a feedback survey, and a semi-structured interview. The first prototype multimedia exam consisted of 40 multimedia questions assessing occupational knowledge in the various AW communities. The prototype was designed and developed by the AW subject matter experts using *Perception for Windows* as the authoring tool. Statistical data from the first data collection led the developers to eliminate two of the items. The revised test consisted 38 questions. The exam questions included a combination of multiple-choice and drop and drag formats. Multimedia components included text, color, graphical images, full-motion video, and audio sound.

The online survey prepared for this study by the authors was used to collect data on the participant's perception and evaluation of the electronic exam process. The first version consisted of nine questions. The revised version consisted of ten questions. The questions were multiple-choice. Participants had the option to write comments.

The semi-structured interview was developed by the development team's psychologist. It was designed to collect data on participants' overall perceptions, their likes or dislikes, and suggestions for improvement.

The electronic exams and surveys were administered on 21-inch display, 200 MHz stand-alone PCs; 14.1-inch display, 1 GHz notebooks; and 10.4-inch display, 800 MHz ViewPad tablet PCs. The runtime program and tests were resident on a CD-ROM. The programs were configured for Microsoft Windows 95 and Internet Explorer 5.5 or higher versions.

**Procedure**

Data were collected during two testing sessions in March and in May, 2002. The tests and interviews took place at the CPRW-11 and Mayport learning centers. Commands have regularly scheduled training days in the learning center and agreed to solicit volunteers. The learning center coordinators assisted in obtaining participants and administering exams and surveys. The development team's psychologist and subject matter experts conducted the semi-structured interviews. Upon arrival at the testing environment, study participants were briefed on the purpose of the study and randomly assigned to a stand-alone, notebook, or tablet PC. After completing the untimed exam, participants filled out the online survey and were debriefed.
RESULTS

Test Data

The electronic exam developers used the same standards for developing the prototypes as the exam developers use in developing the advancement exams used by the Fleet. Current Navy advancement exams are designed to have an overall difficulty level of .54. Acceptable levels range from .20 to .90. Items with values less than .20 are considered too difficult and items with values greater than .90 are considered too easy. Data analysis indicated all of the items in the first prototype were within acceptable range. The overall difficulty level was .56. Two items near the boundaries fell out of acceptable range during the second test session. Overall difficulty level on the second prototype was .60.

Estimates of the reliability of the two prototypes were obtained by computing Cronbach’s alpha. Reliability coefficients of .75 and .76 were obtained for the first and second prototypes respectively.

Feedback Survey Data

Seventy-six (88%) of the participants completed the online survey. Other participants were not able to complete the survey because of their training and flight schedules. Separate analyses of the survey data from the two testing sessions produced similar results. For this paper, the data were pooled and the results of the analysis of the total sample (n=86) are summarized below.

- 97% of the respondents had taken exams for Navy advancement.
- 51% of the respondents had previously taken a computerized test.
- All of the respondents indicated that they had enough computer experience to send email, navigate the Internet, or utilize computer-based training.
- 89% of the respondents felt their computer experience had no effect on their performance.
- 63% of the respondents believed that the prototype took less time, 21% the same amount of time, and 16% more time than a paper-and-pencil exam (p&pe).
- 62% of the respondents believed the electronic exam was less difficult than the p&pe. 8% believed it was more difficult.
- 83% of the participants preferred the computerized exam to the p&pe. 8% preferred the p&pe.
- 80% of the respondents believed that the multimedia exam improves the assessment process. 7% believed that it did not improve the process.
- 96% (n=27) liked having raw scores immediately after completing exams.

Sixty-one (72%) participants provided written comments. In general, the respondents had very positive perceptions of the electronic exam. They enjoyed taking the exam and felt the use of multimedia made the questions easier to understand. Seventeen respondents (28%) described the electronic exam as an improvement over the current system. Sixteen respondents (26%) expressed appreciation for the real world authenticity. Eleven respondents (18%) thought the exams were the way of the future and should be implemented as soon as possible. The reaction to the immediate feedback the exams offered was very positive. Negative comments included remarks that the learning center computers were too slow, some video frame time displays were too long and some of the graphics had poor resolution.

Interview and Observational Data

Participants were observed during test administration to note test-taking behaviors, reactions to the multimedia format, difficulties experienced, and any software or hardware problems encountered. Participants were interviewed or debriefed upon completion of testing.
for additional insights. Debriefing the participant is a basic procedure of usability testing. Debriefing gives participants the opportunity to review their performance and ask questions and it helps the developer understand their errors, difficulties, behaviors and thought processes. Findings from the first testing session include the following.

- Participants preferred a single screen display for each question versus a continuous scroll.
- Participants requested ability to flag questions so they could easily return to them. They also wanted an indication of the questions they were working on and which questions they had completed.
- Participants were "awed" by the tablet technology but did not see the tablets as "Sailor proof" as the desktops. They experienced steep learning curves with both the stylus and the built-in mouse.
- Participants especially liked the multimedia aspects of the exams for their authentic context.
- Participants viewed the exam as an educational tool that not only exposed them to acoustical and visual displays that they cannot learn on paper but also provided immediate feedback on their strengths and weaknesses.

The findings were used to redesign the prototype for the second testing session. Interview data indicated the participants were pleased with the revised prototype. The only design changes recommended included having the audio files end when you proceed to the next question and having an option to enlarge the video displays. There were no negative reactions to the electronic tests. Test takers liked the experience and thought electronic exams took less time, and were easier than paper-and-pencil exams because they did not have to worry about erasures or completely filling the answer spaces.

DISCUSSION

The test taker preference for computer-delivered tests and the positive attitudes toward computerized testing found in this study are consistent with previous findings. Bicanich, Slivinski, Hardwicke & Kapes, (1997) found students preferred Internet delivery to paper-and-pencil versions by a 3-to-1 margin. The majority of studies have found positive attitudes and have indicated that the more experience test-takers have with computers the more positive their attitudes toward computerized testing (Bicanich, Slivinski, Hardwicke & Kapes, 1997; Bocij & Greasley, 1999; Sutton, 1997).

One common concern among test sponsors and stakeholders is that success on computer-based tests requires a high level of computer proficiency. The Sailors who participated in this study use computers daily in the performance of their jobs and are not representative of Navy ratings a whole. Other studies have found computer inexperience did not significantly affect performance on computerized tests (Bicanich, Slivinski, Hardwicke & Kapes, 1997; Wise & Plake, 1989).

The educational benefits of the multimedia exam that the participants expressed are reported in other studies. Bailey (2001) reported that providing more interactivity appears to have a substantial positive effect on learning. Interaction that is cognitively engaging promotes elaborative processing. Multimedia contains more features that are available for extra cognitive processing of information. Test takers take more time to analyze and synthesize the material with prior knowledge which improves learning.

Bocij and Greasley (1999) reported that computer-based techniques may improve overall performance by improving task performance. They found that computer-based tests tend to be completed more quickly than conventional assessments and argued that the assessment technique itself may be acting to enhance the ability of students to focus on questions and recall relevant
information. Interview data and written comments obtained in our study substantiate their argument.

This study was designed to evaluate operator usability and functionality of an electronic exam prototype on various computer delivery systems. All systems performed equally well. Participants liked the pen-based tablet PC but found the built-in mouse and stylus difficult to use.

CONCLUSION

The data suggest that the Sailors in this study are ready to accept the wide-scale introduction of computer-based assessment in place of traditional assessment. Because the communities involved are not necessarily reflective of the Navy as a whole, caution should be exercised in generalizing the results. Usability testing provided a means for improving the electronic exam design and development through quality control processes. Our goal was to understand users and the nature of the task and ensure that human factors in interacting with computer were appropriately addressed in the design. From an HCI perspective, the users interacted with the technology in a manner that was effective, efficient, and satisfying. Standardization of equipment is important in high-stakes testing. The ViewPad Tablet appears to be a viable delivery option; however, issues with the stylus warrant further study. It is recommended that future studies continue to explore the effect that delivery variations might have on test performance. The strong positive responses shown by the participants to the electronic prototypes suggest that the more extensive study planned for Phase II is warranted.

REFERENCES


VALIDITY OF THE ASSESSMENT CENTER FOR FUTURE PROFESSIONAL OFFICERS (ACABO)

Dr. Hubert Annen & Lic. phil. Simon Gutknecht
Lectureship for Military Psychology & Military Pedagogy
Military Academy at the Swiss Federal Institute of Technology
Steinacherstrasse 101 b, CH-8804 Au
hubert.annen@milak.ethz.ch

The tool
The Assessment Center for the Swiss armed forces professional officers has been used since 1992. Until 1995 it served as an appraisal tool which provided important information regarding social behaviour. Because in addition to this, intellectual faculties and practical skills are assessed during academic periods and military services respectively. In 1996 the ACABO became a selection tool (Steiger & Annen, 1997). Students who want to become professional officers must pass the ACABO before they start studying at the Military Academy (MILAK). ACABO is run three times a year with some 20 to 30 candidates.

<table>
<thead>
<tr>
<th>requirements</th>
<th>selection</th>
<th>study course (Military Academy)</th>
<th>objective</th>
</tr>
</thead>
<tbody>
<tr>
<td>qualification for university entrance</td>
<td>militia officer</td>
<td>ACABO 1&lt;sup&gt;st&lt;/sup&gt; year of study course (basic study)</td>
<td>professional officer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Practical training</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt; year of study course (specialised study)</td>
</tr>
</tbody>
</table>

Fig. 1: How to become a Swiss armed forces professional officer

The ACABO is a classical three-day assessment center. The candidates have to deal with reality-based tasks in group discussions, presentations and role plays. Furthermore, they have to take three cognitive tests which assess their analytical and combinatorial thinking. In a written self-portrait they are also required to reflect on their own strengths and weaknesses. During these activities, the candidates are closely observed by several assessors and rated according to specific behavioural dimensions. These criteria are based on the job profile of the professional officer and focus on personality traits and social behaviour (Steiger & Annen, 1997; Annen, 2000).

In order to guarantee a fair and well-based judgement, the assessment follows a process involving several stages. During the perception and assessment stage, observation and judgement must be strictly kept separated. Next, results from individual main or secondary observers are thoroughly discussed after each exercise. And finally, each candidate is again discussed at the observers conference. In cases where the assessment matrix does not produce a homogeneous picture of the candidate, the various results are again examined. The discussion lasts until the observers reach consensus on the profile of the candidate. This systematic approach should keep subjectivity in check and thus minimise typical appraisal errors.
The quality of an appraisal tool mainly depends on how it is accepted by the candidates and on how well it is carried out. Therefore, the ACABO depends, to a large extent, on the competence and the commitment of the assessors, who are either superiors of the candidates and commanding officers of a training course, or militia officers who are professionally involved in personnel appraisal and selection. Both categories are thoroughly trained for and briefed on their task during ACABO.

The procedure is very labour-intensive: An ACABO with 24 candidates requires a staff of 30 persons. Such is only possible because half of the assessors are militia officers doing their regular military service. This situation also exemplifies the huge advantages of the Swiss militia system.

**Evaluation**
At the ACABO, the decisions taken have far reaching consequences for the candidates. Should the candidate fail, he or she would have to wait at least two years before repeating the assessment. This regulation has a strong impact on the career which is then either delayed or even cut short. Therefore, despite its low failure rate of some 15%, this appraisal tool needs regular evaluation in order to guarantee its validity. For this reason, both candidates and assessors have to hand in structured feedback after completion of the assessment center. The results of the evaluation flow into the enhancement of the procedure. In addition, two minor and three longer studies have been carried out in the last two years. Two of them will be shortly presented in the following part.

**Social validity**
Based on the concept of social validity (Schuler, 1998) the candidates have to fill in a questionnaire after undergoing the ACABO. The questions they are required to answer are:

- Do you think that you have been able to demonstrate your strengths?
- How are you feeling – apart from the normal exam stress?
- How realistic do you consider the exercises?

There is also space for notes of criticism and suggestions for improvement. The following table shows that, generally, the results are extremely positive (Annen, 2002).

<table>
<thead>
<tr>
<th>able to show one’s strength</th>
<th>well-being</th>
<th>reality of exercise</th>
</tr>
</thead>
<tbody>
<tr>
<td>completely</td>
<td>feeling well</td>
<td>very realistic</td>
</tr>
<tr>
<td>rather yes</td>
<td>feeling quite well</td>
<td>quite realistic</td>
</tr>
<tr>
<td>rather not</td>
<td>feeling rather unwell</td>
<td>rather unrealistic</td>
</tr>
<tr>
<td>not at all</td>
<td>feeling unwell</td>
<td>unrealistic</td>
</tr>
</tbody>
</table>

According to this study, the participants seem to have enough opportunities to demonstrate their abilities, they do not feel too much under pressure and are not confronted with unrealistic exercises. The result obtained is crucial with regard to participants’ reactions to the AC because Gilliland (1993) claimed that job relatedness
was likely to have the greatest procedural influence on candidates’ perceptions of fairness in the context of personnel selection. Overall, conditions are favourable for the final report to be taken seriously by all persons concerned.

Hophan (2001) examined the extremely positive results thoroughly by analysing the participants’ feedback in a subtly differentiated way and tested several hypotheses against it. They can be brought together in five groups which are shortly reported on as follows:

- **Geographical areas:** The findings revealed no significant differences between German-, French- und Italian-speaking participants with regard to their opinions about the AC with exception of fears and doubts relating to the AC. Reference is made here to the request that the Italian- and French-speaking participants made in their comments to hold the AC introductory session in their mother tongues (and not just in German). As information provided is a crucial factor in reducing the uncertainty of the selection situation (Arvey & Sackett, 1993), it can be concluded that this factor might have contributed to the significant differences between the German-speaking participants and the two “Latin parts”.

- **University degree:** The result showed that participants who had a university degree and those who did not tended to be similar with regard to their opinions about the AC since no significant differences were observed.

- **AC result (passed or failed):** The findings revealed no significant differences between participants who passed the AC and those who failed. This result appear to be in line with Teel & DuBois’ (1983) investigation which found that both high- and low-scorers reacted positively to the AC programme.

- **The participants’ age:** As no significant difference could be found, evidence is provided that the participants’ age range (21 to 34 years) does not tend to influence participants’ opinions about the AC.

- **The officers’ rank:** This question is important as one might have expected different responses from lieutenants, first lieutenants and captains as all of them have had out different practical experience and therefore different expectations about exercises and contents. Yet no significant differences could be found in this respect.

Altogether it can be said that the high social validity of ACABO stands up well to closer scrutiny.

**Incremental validity**

With structural equation modelling, Gutknecht (2001) analysed the various tools used for the selection of professional officers and examined their validity concerning study success and vocational success. The advantage of this procedure is that it uses latent variables such as cognitive skills. The impact of the variables are simultaneously tested. The two criteria study and vocational success could not be tested in one model due to lack of data. Therefore, the potential determinants of study success were examined in a first study (n=76). The following items were taken into consideration:

- school performance,
- the cognitive skills such as measured by the cognitive tests during ACABO
- the results from ACABO („social competencies“).
Based on the results of the confirmatory factor analysis, the variables shown in the model can be considered construct valid. As to “study success”, the hypothesis confirms that the variables “school performance” as well as “cognitive skills” predict “study success” accurately (p < .05). These two variables disclose considerable additional information regarding “study success”. As expected, the assessment of “social competencies” from ACABO has a low predictive value for the variable “study success”.

A second study (n=104) examined the possible predictors for “vocational success” and their interdependence. It took into consideration the factors “social competencies” which were thought to be most likely to produce predictions about the “vocational success”. It furthermore considered “school performance”, “study success” during the first and the third year at MILAK as well as the record during the year of “practical training”. The measurement models for each latent construct were good, that means they are construct valid. In a first partial study, the “vocational success” was then defined by the results from common job appraisals carried out by the superiors. In a second partial study, a prestigious indicator (belonging to general staff) was additionally taken into account for the operationality of the factor “vocational success”. The following figure shows the predictors in chronological order, the brackets contain the beta-coefficient of the second partial study.
As the figure shows, the highest direct influence is derived from the variable “study success” during the third and last year of study, and this independently of the operationality of the outer criterion. The direct predictive value of the factor “social competencies” is rather small. If the total effect of the variables are taken into consideration, then the coefficient rises to .47, but still remains way below the total effects of the variable “study success” during the first and third year of study. As a result, the hypothesis saying that the “social competencies” have the highest predictive value concerning “vocational success” has to be rejected.

It is clear that the influence of the variable “social competencies” is mediated by the appraisal resulting from the “practical training”. If this variable is removed from the model – as it plays only a minor role in the whole selection process – the direct influence of the “social competencies” under the condition “vocational success with prestige factor” dramatically increases (.41*), but it still remains below the values of “study success” in the third and last year of study (.43*). Based on this last findings it can be said that ACABO contributes a certain amount of additional prognostic validity concerning “vocational success”. Whether this is enough, remains to be examined in the light of a cost-benefit analysis.
Conclusion and outlook
ACABO is the appraisal tool of the Swiss Armed Forces, which has been shown to be the most scientifically accurate. Its quality is also praised by those working in the “field”. It is for these reasons that recent selection and appraisal tools have often been modelled on ACABO, such as the Assessment Center for future NCOs (ACABUO), the new selection procedures for militia officers (Boss, 2001) and the new regulation for the qualification of militia cadre in practical services and in refresher courses (Annen, 2000).

The Swiss Armed Forces are in the middle of a comprehensive reform. The training and careers of future militia officers will change dramatically. The need for contract and professional military personnel will increase substantially. Given this background, the question is whether the form and timing of ACABO as well as the resources used are still adequate. Although the practical and scientific benefits of ACABO have been proven, this tool is not a given. One must anticipate changes and be open to new forms. Scientific investigation has yielded important data regarding this issue. Such data can be considered to be the foundation for the development and application of comparable diagnostic instruments.

References


REDUCING ADVANCED INDIVIDUAL TRAINING (AIT) TO LESS THAN 26 WEEKS\textsuperscript{15}

Kelly S. Ervin, Ph.D.
Research Psychologist
U.S. Army Research Institute for the Behavioral and Social Sciences
Ervink@ari.army.mil

Introduction and Overview

Advanced Individual Training (AIT) is the US Army's first opportunity to provide job-specific training to newly enlisted soldiers. AIT may run for 26 weeks or longer, depending on the Military Occupational Specialty (MOS). The US Army Training and Doctrine Command (HQ TRADOC) is concerned that this extended period of training, before actual job assignment, may cause new soldiers to experience skill decay and knowledge retention difficulties. The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI), Occupational Analysis Office (OAO) undertook a study to determine if graduates of lengthy AITs are experiencing these problems.\textsuperscript{16}

Skill and knowledge decay is a loss of trained or acquired skills after periods of nonuse, or infrequent opportunities to practice or perform the acquired skills (Ford, Quiñones, Sego, & Speer Sorra, 1992). Because of this, there is merit to HQ TRADOC’s concerns about the lengthy period between when an individual joins the Army, completes nine weeks of basic combat training, 26 weeks of AIT, and finally progresses to the first unit assignment. The soldier must go through approximately nine months of training before he or she begins the work of the Army. During that nine-month period, skills trained and knowledge learned in the early portions of the training may be forgotten. For example, skill decay is seen as a particular problem for Reserve personnel in the military (Wisher, Sabol, Hillel, & Kern, 1991).

There is consensus in the human performance literature that a core set of seven factors exists which influence skill decay and knowledge retention. Factor 1: Length of retention interval, the time in which acquired skills and knowledge are not used or exercised. In general, longer periods of nonuse are associated with greater decay (Naylor & Briggs, 1961; Annett, 1979). Factor 2: Degree of overlearning, providing students additional training over and beyond what is considered sufficient for skill proficiency. Overlearning strengthens the bond between stimuli and memory and increases skill and knowledge retention (Schendel & Hagman, 1982; Driskall & Sals, 1992). Factor 3: Task characteristics, there are six groupings of task characteristics that affect the difficulty of task acquisition and retention: physical versus cognitive, closed-looped versus open-

\textsuperscript{15} All statements expressed in this paper are those of the author and do not necessarily reflect the official opinions or policies of the U.S. Army Research Institute or the Department of the Army.

\textsuperscript{16} Human Resources Research Organization (HumRRO) of Alexandria, Virginia was awarded the contract (DASW01-98-0047) to assist ARI-OAO in conducting this study.

44\textsuperscript{th} Annual Conference of the International Military Testing Association
Ottawa, Canada, 22-24 October 2002
looped\textsuperscript{17} (Naylor & Briggs, 1961), natural versus artificial, integrated versus nonintegrated (Annett, 1979) and level of difficulty (Mumford, Weeks, Harding, & Fleishman, 1987). Factor 4: Methods of testing for original learning & retention, or the types of measurements used to assess skill and knowledge retention. That is, retention scores may be a function of the type of test used to measure the retention. For example, there is a difference between recall and recognition testing and the scores on a recall test may be lower than scores on a recognition test. Meaning, one’s ability to recall a skill does not necessarily translate into one’s ability to remember how to perform that skill. Factor 5: Conditions of retrieval, which facilitate skill and knowledge retention, are rooted in the \textit{encoding specificity principle}. This principle states that retention will be maximized if the retention assessment matches, as closely as possible, the environment in which learning first took place (Tulving, 1983). This is because environmental cues will facilitate memory and retrieval of learned information (Schab, 1990; Cann & Ross, 1989). Factor 6: Instructional & training methods, different methods can influence skill and knowledge retention. For example, on-the-job, lecture, computer-assisted, machine simulators, and team training. That is, some pedagogical methods, or a combination of methods, facilitate retention more than others. Factor 7: Individual differences, there is some evidence that individuals with high learning abilities retain more knowledge for longer periods than individuals with low learning abilities (Farr, 1987).

Much of the human performance literature is filled with articles on individual aspects of human performance and the factors that facilitate or inhibit optimum performance. In an effort to combine these fragments of data into a composite of knowledge, Arthur, Bennett, Stanush, and McNelly (1998) conducted a meta-analysis on over 200 empirical articles on human performance factors. Their work revealed that skill and knowledge retention decreased as time between training and performance increased and the decay associated with physical and natural tasks is less than the decay associated with cognitive and artificial tasks. They also found that using recognition measures to test skill and knowledge retention resulted in less evidence of decay than when using recall measures. However, in contrast to much of the evidence found in the human performance literature, Arthur and his colleagues found that closed-looped tasks were associated with less skill decay than open-looped tasks (see footnote #3). These seven factors can be considered as two categories of influences: methodological and task related. Methodological influences are those factors that can be modified in the training instruction and environment to reduce skill decay and enhance knowledge retention (overlearning, conditions of retrieval, evaluation criteria & method of testing). On the other hand, task related factors are those that are inherent characteristics of the task and usually cannot be modified by the trainer.

The major premise of this study, that AIT's 26 weeks or longer presents a significant time delay between training and application of that training (i.e., performance) is a concern for methodological influences. That is, time between training and application of that training is in control of the trainers and job supervisors while training developers and instructors

\textsuperscript{17} Closed-looped tasks are tasks that involve distinct responses and that have an exact beginning and end. Open-looped tasks are those tasks that involve continuous, repeated responses without an exact beginning and end.
determine the training methods used. Indeed, the length of the retention interval is a powerful predictor of retention and is a factor that is mediated by other influences (e.g., overlearning, task characteristics, methods of testing, etc.). This point has particular implications for the idea that skill decay and knowledge retention is adversely affected by lengthy AIT courses. If AIT graduates demonstrate skill decay and difficulty in retaining acquired knowledge, it may not be because the course was too long, but in fact, may be a combination of length and any of the other factors identified as affecting retention. The question becomes, how best to train AIT students so that the skills and knowledge acquired are not lost.

The Study

ARI-OAO is engaged in examining various aspects of AIT and soldier performance to determine pedagogical methods used in AIT training, the specific types of tasks that are being trained in AIT and what tasks AIT graduates are expected to perform during their first unit assignment. In addition, data are being collected that illustrate which tasks are being performed to “standard” and which tasks, if any, could be trained by performance support systems (e.g., distance learning (DL) or on the job training (OJT)) after AIT. Tasks not being performed to “standard” are seen as evidence of skill decay and knowledge deficiency and for those tasks, information will be collected concerning the perceived reasons (e.g., no opportunity to practice or use skill, working on equipment without appropriate training, etc.) for the performance deficiencies.

The data collection effort for this study is focusing on three MOS representing three different schools: The U.S. Army Intelligence Center which trains Military Intelligence Systems Maintainer/Integrators (MOS 33W) at Ft. Huachuca, Arizona (course length 43 weeks), the U.S. Army Aviation Logistics School which trains Armament/Electrical Avionic Systems Repairers (MOS 68S) located at Ft. Eustis, Virginia (course length 29 weeks) and the U.S. Army Signal Center which trains Air Traffic Control Systems, Subsystems & Equipment Repairer (MOS 35D) at Ft. Lee, Virginia (course length 34 weeks).

We began with current school-approved task lists for the three MOS being studied. Developed by Subject Matter Experts (SMEs), the MOS task lists are comprehensive inventories of individual tasks a job incumbent may have to perform. The task lists are being studied by additional SMEs and will be revised, if necessary, to ensure that the current tasks trained are those that are still needed for optimum job performance and to include additional tasks that are not on the original list. Along with SMEs, ARI will interview training developers, MOS instructors, and current students to discover what pedagogical methods are used as a function of task requirements. This study is currently in progress and we have completed interviews at the Military Intelligence School (MOS 33W). The results of these interviews are described in the Interviews section of this paper.

Using revised and updated task lists we will then collect data from AIT graduates and their supervisors by means of a computer-assisted questionnaire. ARI-OAO will develop
automated questionnaires for each MOS by using the Automated Survey Generator (AUTOGEN) system, a software program that enables users to develop questionnaires, collect data, import data into statistical software programs (e.g., SPSS) and generate reports. Respondents will indicate which tasks they perform, how often they perform them, to what extent they perform tasks to “standard,” at what point, during AIT, did they acquire the knowledge to perform the task (e.g., beginning, middle or end), their impressions of their AIT course and their opinions on which skills suffer from time delays between training and expected performance. Supervisors of AIT graduates will be asked similar questions and will also be asked to provide their opinions on whether or not AIT has prepared the soldiers to perform their jobs, on which task skills do they notice skill or knowledge deficiency and what, if anything, could be done to minimize decay and retention problems.

Interviews

We interviewed current students, instructors, training developers, and SMEs (e.g., training division chiefs) in separate groups. We toured several classrooms and in some instances, were able to observe actual training in progress. There are currently 835 soldiers trained as 33Ws throughout the Army and the purpose of the 33W AIT course is to produce journeymen capable of troubleshooting most types of electronic equipment.

Three themes emerged from the interview data in response to our questions concerning length of AIT for MOS 33W, tasks trained, and reasons for possible skill/knowledge decay associated with this MOS.

Encountering equipment in the field that was not including in AIT. The current students were generally satisfied with their AIT experience and expected to encounter electronic equipment that is different from what they were exposed to in the classroom. They felt confident that AIT provided them with the critical thinking skills and practical knowledge needed to repair a variety of electronic systems. The students understood that the goal of the course was to train them to understand systems of electronics and that knowledge of specific equipment was of little consequence. In fact, the objective of the course is to ensure that graduates, who will encounter one-of-a-kind equipment in the field, are able to repair that piece of equipment down to a component level. The students also felt that prior experience with electronic equipment before AIT was a major factor for success in the course and subsequently in the field since one may encounter unfamiliar equipment. The students also indicated that they knew, once in their first unit assignment, they could engage in correspondence courses and/or individual time with their job supervisor to practice skills not being used.

Unit assignments may not match MOS training. The students, instructors, and SMEs agreed that one of the more serious issues relating to skill and knowledge decay is that many graduates of MOS 33W may not be assigned to units that require their particular skills. In many instances, 33Ws are being required to work on tasks not trained and are subjected to non-MOS duty (e.g., mowing grass, painting buildings, etc.). Related to this was the concern that 33Ws suffer from skill and knowledge decay because of a high
reliance on contractors hired to do electronics repair. The use of contractors was given as a particular concern for facilitating low morale, frustration, and disillusionment among the soldiers.

Length of AIT. The instructors and SMEs were quite adamant that the current length of the course was adequate (26 weeks) and that shortening the course would negatively affect students’ ability to perform tasks to “standard” once they are in their unit assignments. There was consensus that the length of the course did not need to be shortened but time spent on certain topics needed to be redistributed. Too much time was spent on certain topics (e.g., fiber optics) and not enough time was spent on other topics (e.g., soldering). All agreed that the course should not be any longer than 26 weeks. The instructors, training developers, and SMEs felt that many of the young individuals currently joining the Army lack critical thinking skills and show no desire to acquire them. In the field, they notice that the younger soldiers do not want to take the time to read training manuals and want to be told “what to do” instead of thinking and solving problems for themselves. In addition, they felt that the only way the AIT could be successful at less than 26 weeks was to require a prerequisite of at least 2-years of college level electronics training.

Future Work

Interviews with current AIT students, training developers and SMEs at MOS 68S and 35D are scheduled for the near future and the results of the interviews will be compiled with the interview data from MOS 33W to assist us in developing appropriate questions for the AUTOGEN questionnaire. As the raw data are received from the respondents, we will formulate hypotheses about trends and patterns concerning any evidence of skill decay and knowledge retention. Data will also be collected on which tasks are being performed to “standard” and which tasks, if any, could be trained by performance support systems such as assignment orientated training (AOT) after completion of AIT. To support the related AOT initiative a statistical analysis program in AUTOGEN will identify clusters of Skill Level 1 soldiers based on unique task performance as well as identifying core tasks that are common to soldiers across all clusters.

The results of this study are designed to provide a general set of procedures for evaluating AIT. This set of procedures will provide training developers and instructors a set of guidelines to assist them in ensuring that AIT provides the knowledge, skills, and abilities needed by soldiers when they reach their units of assignment. In addition, the procedures will provide guidance on pedagogical methods that facilitate skill and knowledge retention. This study’s projected completion date is February 2003.

References


ORDER EFFECTS OF TASK INVENTORY DESIGN

Natalie Collins
Directorate of Strategic Personnel Planning and Research
Department of Defence, ACT, Australia 2600
Natalie.Collins1@defence.gov.au

Patricia Brown, Ph.D.
Lecturer in Psychology, Centre for Applied Psychology
University of Canberra, ACT, Australia 2601
Patricia.Brown@canberra.edu.au

ABSTRACT

Precise job information is a vital resource for managers, with the task inventory approach to Job Analysis/Occupational Analysis being a valuable tool in collecting reliable and valid job information. However, research on the order of tasks in the inventory has been limited, with the alphabetised by verb approach frequently advocated as the most valid and easiest for respondents to complete. The current experimental study was designed to determine whether a different set of responses are produced when the same set of tasks are ordered using scripts, compared to when alphabetised by verb. A between groups design was applied in the current study where participants were matched based on their unit function (for example, participants from training establishments were matched), and assigned to either the script based or alphabetised by verb task inventory. Using a sample of 434 other ranks (ORs) in the Explosive Ordnance specialisation, analysis of variance results indicated that junior ORs in the alphabetised group took significantly less time to complete the task inventory compared to both junior ORs in the script group, and senior ORs in the alphabetised group. It was also found that job incumbents in the script group ticked significantly more tasks than those in the alphabetised group, and that senior ORs ticked significantly more tasks than their junior OR colleagues. Results are discussed with regard to cognitive processing theory, and the subsequent implications of the current findings on the design of task inventories within the Australian Defence Force.
INTRODUCTION

The task inventory approach has been used for almost 30 years within the Australian Defence Force (ADF), however, it remains questionable as to whether the order of tasks can influence job incumbents’ responses. As part of reviewing the way the ADF conducts occupational analysis (OA), this study examined “script” and “alphabetised by verb” task orderings. The task inventory approach to OA is a systematic method of collecting job information using a task inventory, and as Melching (1973, p. 3) outlines, “a task inventory is a list of appropriate duty and task statements covering the tasks performed by workers in an occupational area.” There are many advantages in collecting job information using a task inventory. The questionnaire is easily self-administered and task information about the work “actually” performed can be collected from a number of geographically dispersed work areas. Further to this, the data is quantifiable, and can be stored, manipulated, analysed and reported by computer (AF Occupational Measurement Squadron, 1998; Ash, 1988; Christal & Weissmuller, 1988; DDOA & Staff, 2001; Gael, 1983; Melching). Clearly the task inventory approach to OA is a very powerful tool to provide management with high quality job information.

Current Approach to Task Ordering

Obviously the crucial component of a task inventory is the tasks that underpin the list. Given that task inventories can contain 500 tasks or more (Gael, 1983), it is important to order tasks in a manner that allows the analyst to capture accurate and reliable job information, and that minimises the amount of effort required by the job incumbent to complete the questionnaire. The usual approach taken when ordering tasks is to list them alphabetically by verb under each duty area heading. The rationale given for this ordering is that the duty area headings can be used as a guide for the job incumbent, and they can be asked to quickly scan duty areas that are not part of their responsibility, but pay close attention to areas that are (Christal & Weissmuller, 1988). Furthermore, the following reasons are often cited as to why tasks should be alphabetised by verb:

?? it can assist the analyst identify and remove duplicate tasks;
?? it can help the job incumbent recall tasks that are missed (for example, the incumbent can scan all the ”repair” tasks to ensure that all the repair tasks he/she performs are in the task inventory);
?? it can shorten the job incumbent’s reading time (Christal & Weissmuller; McCormick, 1979; Melching, 1973); and
?? it is asserted that a more accurate measure of the relative time spent on tasks is obtained, “as each task is considered as a discreet action” (Boyce & Garman, 1995, p. 148).

It would therefore seem that reasons as to why tasks should be ordered alphabetically by verb could be divided into two groups. Those concerned with the development of the inventory, and those with the effects of inventory administration.

Other Approaches to Task Ordering

A study by Coover (1990), using university students evaluated response times to each task within a “student” task inventory using four different task orders. The tasks were
ordered in one of the following ways, by verb similarity, noun similarity, meaning of task statement similarity, or script. Coovert concluded that response times for individual task statements were fastest when the task inventory was organised using scripts. Furthermore, Getkate (1994) found that scripts provide job incumbents with a meaningful interpretation of the way they experience work. Scripts are a “predetermined, stereotyped sequence of actions that define a well known situation” (Schank & Abelson, 1977, p. 41). Thus, scripts are a type of schema that are structured with events causally linked together, and apply to mundane or routine events (Medin & Ross, 1997). It seems that scripts promote cognitive efficiency by providing people with a framework to economically organise and process information.

What is pertinent to this study, however, is that script-theoretic and job-analytic views of work are very similar (Getkate, 1994). Given the layout of a task inventory is organised in a way similar to a script, and assuming all job incumbents have similar work scripts, it would seem logical that a script ordered task inventory would be easier for a job incumbent to complete compared to one alphabetised by verb. Nevertheless, the following issues need to be considered:

?? The major limitation of the research mentioned above is that work scripts have only been examined for job incumbents at a similar rank or level.

?? Scripts must be written from the view of one particular role (Schank & Abelson, 1977), and consequently job incumbents at different ranks or levels could adopt different roles within an organisation, and have differing work scripts (e.g., “managers” versus “hands-on”).

?? If different work scripts do exist, any benefits of the script based inventory could potentially be lost, as scripts are sequential in nature and that each component of the script must be completed (Schank & Abelson). If previous actions are not completed satisfactorily, hitches must be dealt with (Schank & Abelson), hence, any heightened levels of activation in the script’s action nodes are lost.

?? The gap-filling phenomenon, whereby items not originally presented but consistent with the script are falsely recognised. That is, when people are presented with a series of items in a script format, when asked to identify the original items, they report “extra” items that are consistent with the script but not presented as part of the original stimulus (Bower, Black & Turner, 1979; Graesser, Woll, Kowalski & Smith, 1980).

Clearly then there are other theoretical and practical consequences that may counteract any benefit gained by using a script ordered task inventory.

The Present Study

The current study was designed with two primary aims. Firstly, it aimed to assess whether job incumbents found it easier to complete a task inventory that was alphabetised by verb or one using scripts. Secondly, the study endeavored to determine if the ordering of tasks in a task inventory affected the job information provided. The study was also

---

18 A script task order refers to ordering tasks in a sequence that they are preformed at work.

19 For a complete review on a script-theoretic and job-analytic view of work, see Getkate (1994).
designed to assess whether the rank (or role associated with that rank) of a job incumbent impacted on the above two aims.

METHOD

Participants

The participants were 434 other ranks (ORs) (404 males, 17 females, and 13 unspecified gender) working in the Explosive Ordnance (EO) specialisation who were part of a larger OA being conducted by the Department of Defence. These participants had a mean age of 31 years and two months ($SD = 7.50$ years, range 18 years to 53 years). The three ADF services were represented in this study by 198 Navy, 153 Air Force, and 83 Army participants, who on average had been with their service for 11 years ($SD = 7.19$ years, range less than one year to 35 years). In terms of rank, 275 participants were junior ORs (Private to Corporal equivalents), and 159 senior ORs (Sergeant to Warrant Officer equivalents).

Materials

The material used in this study was a self-report questionnaire, of which there were two versions. The two versions of the questionnaire contained different task inventories; the control version of the questionnaire had a task inventory ordered alphabetically by verb, whilst the experimental version of the questionnaire used a task inventory where tasks were grouped in scripts. In all other respects the task inventories were exactly the same, as they contained the same duty headings (in the same order), and tasks in both inventories were worded exactly the same way. To demonstrate the difference between the two task orderings used in this study an example of the first five tasks under the duty heading "General Technical" from each inventory is shown in Table 1.

Table 1: Examples of Task Ordering in Each Type of Inventory – “General Technical” Duty Area

<table>
<thead>
<tr>
<th>Alphabetised by Verb</th>
<th>Script Based</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Activate fixed fire fighting appliances</td>
<td>1. Raise fire alarms</td>
</tr>
<tr>
<td>2. Administer advanced EO related &quot;life support&quot; procedures such as Defibrillation</td>
<td>2. Attack fires with first aid appliances (such as fire extinguisher, fire blanket)</td>
</tr>
<tr>
<td>3. Administer emergency EO related first aid</td>
<td>3. Activate fixed fire fighting appliances</td>
</tr>
<tr>
<td>4. Administer the controls for prohibited articles within storage facilities</td>
<td>4. Muster fire attack parties</td>
</tr>
<tr>
<td>5. Approve EO security SOPs</td>
<td>5. Participate in Fire Attack Party activities</td>
</tr>
</tbody>
</table>

Development of the Questionnaire

Three subject matter expert (SME) panels (each consisting of eight different experienced senior members of the EO specialisation) were conducted to develop the task inventory section of the questionnaire. During these panels, SMEs were asked to order the tasks in
the sequence in which the work would be performed for an activity (i.e. to construct a script for each logical grouping of tasks). At times there was disagreement as to what order the tasks would be performed, but in these situations, the order according to the majority of members on the SME panel was included in the inventory. The task inventory developed at the end of the SME panels was then distributed widely for other members within the EO specialisation to make comment on. During this consultation junior members within the EO specialisation were targeted to ensure that the task inventory accurately reflected the work that they perform. At this point a total of 569 EO tasks had been identified (under 17 duty area headings), which consequently was the total number of tasks used in the final questionnaire. The researcher then took the task inventory that had been developed with tasks grouped in scripts, and alphabetised by verb the tasks under each of the duty areas.

**Dependent Measures**

*The four dependent measures used in this study were as follows.*

- **Ease of Task Identification** – this item asked participants to indicate how easy it was to identify the tasks they performed, and used a five-point scale that ranged from *very difficult* (1) to *very easy* (5).
- **Time Taken to Reach End of Task Inventory** – this item asked respondents to indicate the amount of time (in minutes) it had taken them to complete the task inventory.
- **Number of Tasks Ticked** – this item collected information on the number of tasks ticked by participants.
- **Relative Time Spent on Tasks** – this item collected information on the relative time participants spent on tasks, and used a seven-point scale that ranged from *a very small amount* (1) to *a very large amount* (7).

**Procedure**

Given that past experience has shown that members of units often sit down and complete task inventories together, a different assignment method was adopted for units with more than one EO participant and for units where there was only one EO participant. Therefore in units where there were at least two EO personnel the entire unit was allocated to either the control or experimental group, whereas individual EO personnel were randomly assigned to either group. Where there was more than one EO member at a unit, the units were matched based on their function such as, training, and these matched units were assigned to either the control or experimental condition.

The survey was conducted from 19th February 2002 to 5th May 2002. An overall response rate of 50.5% was achieved for the survey, whilst on a task inventory basis a 47.5% response rate was achieved for the alphabetised version and 53.3% for the script based inventory.

**RESULTS**

Basic data screening was conducted, with 39 respondents being removed from the initial sample as they ticked less than five tasks in the task inventory, and a further two
respondents as they ticked all tasks in the task inventory. This left 393 valid responses for further analysis.

Equivalence of Groups – Demographic Profile

As total random assignment was not possible in the current study, it is important to demonstrate the equivalence of groups based on demographic variables. Of the 393 valid responses, 181 were from the alphabetised group, and the remaining 212 from the script group. To establish whether the participants in these groups were similar, a series of descriptive statistics were calculated for each group. Independent groups t-tests and Chi-Square for relatedness or independence tests were then conducted to determine if the demographic profile of the alphabetised and script groups was statistically different.

A series of three independent groups t-tests were conducted for the demographic items of age, time in service, and time in position. For age there was no significant difference between the alphabetised and script groups \( t(386) = -.09, \ p = .932 \). No significant difference was also found for time in service \( t(383) = .53, \ p = .596 \), and for time in position \( t(380) = -.41, \ p = .685 \).

As the variables service, rank, Explosive Ordnance (EO) role, and gender are categorical variables, the Chi-Square for relatedness or independence test was used to evaluate equivalence between the alphabetised and script groups. There was no statistically significant difference in terms of the alphabetised and script groups based on service, \( \chi^2(2, N = 393) = .83, \ p = .661 \). Furthermore no differences were found for the groups based on rank, \( \chi^2(5, N = 393) = 6.40, \ p = .269 \), EO role \( \chi^2(9, N = 391) = 7.27, \ p = .609 \), or gender \( \chi^2(1, N = 382) = .17, \ p = .679 \).

Overall, the results of the statistical tests on the demographic items indicate that the alphabetised and script groups are equivalent. This means that despite the group assignment method used, the alphabetised and script groups are matched. Furthermore, these demographic items should not be confounding factors on further analyses.

Univariate Analysis of Variance (ANOVAs)

A series of 2 (rank – junior ORs versus senior ORs) x 2 (group – alphabetised versus script) factorial ANOVAs were performed on the dependent variables time taken to complete, ease of task identification, number of tasks ticked, and relative time spent on tasks. The original rank variable was reduced to two levels as the cell sizes of the six level rank variable were too small to be used in ANOVAs. The original rank variable was combined in the following way to produce the two groupings of; junior ORs (Private to Corporal equivalents), and senior ORs (Sergeant to Warrant Officer Class 1 equivalents).

Time Taken to Complete

A 2 (rank – junior ORs versus senior ORs) x 2 (group – alphabetised versus script) factorial ANOVA was performed using the dependent variable time taken to complete. Junior ORs in the alphabetised group reported taking the least amount of time to complete the task inventory \( M = 41.59 \) minutes, \( SD = 26.11 \) minutes), whilst senior ORs...
in the alphabetised group reported taking the longest amount of time to complete the task inventory \((M = 60.38\) minutes, \(SD = 44.80\) minutes).

The assumption of normality for ANOVA was violated by the variable time taken to complete, hence, this variable was logarithmically transformed. The ANOVA indicated that there was not a significant effect for group, \(F(1, 327) = 1.62,\) partial \(\eta^2 = .005,\) power = .25, \(p = .203,\) nor for rank, \(F(1, 327) = 2.77,\) partial \(\eta^2 = .008,\) power = .38, \(p = .097.\)

However, there was a significant interaction between group and rank, \(F(1, 327) = 5.30,\) partial \(\eta^2 = .016,\) power = .63, \(p = .022.\)

To follow up the significant interaction, a series of six pairwise comparisons were conducted. Of the six independent groups t-tests conducted there were only two comparisons that were statistically significant at an alpha level of \(.008.\)

Junior ORs in the alphabetised group took significantly less time to complete the task inventory than junior ORs in the script group \(t(203) = -2.99,\) \(p = .003.\)

Junior ORs in the alphabetised group also took significantly less time to complete the task inventory than senior ORs in the alphabetised group \(t(146) = -2.73,\) \(p = .007.\)

### Ease of Task Identification

A 2 (rank – junior ORs versus senior ORs) x 2 (group – alphabetised versus script) factorial ANOVA was also performed on the dependent variable ease of task identification. The reported means in all cells for ease of task identification were around the 2.6 mark, which indicates that most respondents found it relatively easy to identify the tasks they perform in the inventory. Table 2 presents a summary of the ANOVA, and as can be seen from this table, no significant effects were found.

<table>
<thead>
<tr>
<th></th>
<th>MS</th>
<th>df</th>
<th>(F)</th>
<th>power</th>
<th>effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>.000</td>
<td>1</td>
<td>.00 ns</td>
<td>.05</td>
<td>.000</td>
</tr>
<tr>
<td>Rank</td>
<td>.068</td>
<td>1</td>
<td>.09 ns</td>
<td>.06</td>
<td>.000</td>
</tr>
<tr>
<td>Group x rank</td>
<td>.018</td>
<td>1</td>
<td>.02 ns</td>
<td>.05</td>
<td>.000</td>
</tr>
<tr>
<td>Between group error</td>
<td>.738</td>
<td>317</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(ns =\) not statistically significant at an alpha level of \(.05.\)

### Number of Tasks Ticked

A 2 (rank – junior ORs versus senior ORs) x 2 (group – alphabetised versus script) factorial ANOVA was performed using the dependent variable number of tasks ticked. Junior ORs in the alphabetised group reported the fewest number of tasks ticked \((M = 61.94\) tasks, \(SD = 59.04\) tasks), whilst senior ORs in the script group reported the most number of tasks ticked \((M = 161.36\) tasks, \(SD = 115.91\) tasks).

\(2^0\) Similar results were also obtained using the non transformed data, i.e., a significant interaction between group and rank, and a non-significant main effect for group and rank.

\(2^1\) When follow up comparisons were performed on the non transformed data the same two significant differences were obtained.
The assumption of normality for the ANOVA was violated by the variable number of tasks ticked, hence, this variable was logarithmically transformed. The ANOVA indicated that there was a significant effect for group, $F(1, 389) = 5.95$, partial $\eta^2 = .015$, power = .68, $p = .015$. Therefore as there are only two levels of the independent variable group, examination of the overall means suggest that, those participants in the script group ticked significantly more tasks ($M = 112.62$ tasks, $SD = 105.93$ tasks) than participants in the alphabetised group ($M = 81.85$ tasks, $SD = 78.76$ tasks). There was also a significant effect for rank, $F(1, 389) = 44.77$, partial $\eta^2 = .103$, power = 1.00, $p = .000$. Therefore as there are only two levels of the independent variable rank, examination of the overall means suggest that, senior ORs ticked significantly more tasks ($M = 137.16$ tasks, $SD = 108.51$ tasks) than junior ORs ($M = 74.29$ tasks, $SD = 77.37$ tasks). There was not a significant interaction between group and rank, $F(1, 389) = 1.54$, partial $\eta^2 = .004$, power = .27, $p = .216$.

**Relative Time Spent on Tasks**

The final 2 (rank – junior ORs versus senior ORs) x 2 (group – alphabetised versus script) factorial ANOVA was conducted on the variable average relative time spent on tasks. Using the relative time spent data collected, an average relative time spent score was calculated for each respondent. A respondent’s average relative time spent score was calculated by summing the relative time spent score indicated for each task, and then dividing the overall score by the number of tasks ticked. The reported means for average relative time spent were around the 2.7 to 3.0 mark, which indicates that most respondents spent a below average amount of time on tasks.

The ANOVA indicated that there was not a significant effect for group, $F(1, 389) = .53$, partial $\eta^2 = .001$, power = .11, $p = .467$, nor for rank, $F(1, 389) = 2.371$, partial $\eta^2 = .006$, power = .336, $p = .124$. Furthermore, there was not a significant interaction between group and rank, $F(1, 389) = 1.12$, partial $\eta^2 = .003$, power = .18, $p = .290$.

**DISCUSSION**

*Order Effect Based on Time Taken and Number of Tasks Ticked*

Independent of inventory type, junior ORs ticked significantly fewer tasks than their senior OR colleagues, therefore it seems reasonable to expect that they would take less time to complete the task inventory. However, there was no difference between the time it took junior ORs in the alphabetised group and senior ORs in the script group to complete the task inventory. Nevertheless, the time taken results did indicate that junior ORs in the alphabetised group took significantly less time to complete the inventory compared to junior ORs in the script group and senior ORs in the alphabetised group. Clearly when the time taken to complete the task inventory, and number of tasks ticked results are considered together, a task ordering effect becomes evident.

\textsuperscript{22} Means and standard deviations presented are those prior to transformation.
It appears that the senior OR rank grouping benefited the most from the script based inventory. It is acknowledged that overall, job incumbents in the alphabetised group ticked significantly fewer tasks than those in the script group, and consequently should have taken less time to complete the inventory. However, as this was not the case, the results can be explained as follows. If most of the scripts presented in the script based inventory were inconsistent with the actual work scripts of junior ORs, any activation of nodes was minimal. Therefore the level of node activation of junior ORs in the script group was probably comparable to that of junior ORs in the alphabetised group, as verb schemas only have a confined spread of activation to small number of nodes (Wickelgren, 1979). Consequently it took junior ORs in the script group significantly longer to complete the task inventory compared to junior ORs in the alphabetised group because they had ticked more tasks and did not have any cognitive processing advantage. However, a time difference was not observed for senior ORs, although senior ORs in the script group ticked more tasks than senior ORs in the alphabetised group. The failure to identify a time effect for senior ORs could be attributed to cognitive processing efficiencies from scripts negating any time savings achieved by the alphabetised group ticking fewer tasks. Furthermore, senior ORs in the script group benefited from cognitive processing efficiencies associated with task information presented in role related scripts, whilst junior ORs in the alphabetised group achieved time savings from ticking significantly fewer tasks. Therefore there was no significant difference in the amount of time it took senior ORs in the script group and junior ORs in the alphabetised group to complete the task inventory.

Nevertheless, some of the scripts in the script based inventory must have been relevant to junior ORs, as on the whole, job incumbents in the script group ticked significantly more tasks than those in the alphabetised group. This overall effect can be explained in two different ways. Firstly, there is the possibility that the script based inventory was more “psychologically meaningful” and hence, actually elicited job information that provided a better representation of the work performed (Getkate, 1994). This explanation is plausible as the result obtained for senior ORs is consistent with other research findings on homogenous groups (Coovert, 1990; Getkate). Furthermore, even if junior ORs are a heterogeneous group, the script based inventory would be “psychologically meaningful” in certain areas where common work scripts occur. Therefore junior ORs in the script group would tick significantly more tasks than junior ORs in the alphabetised group. An alternative explanation is that the script based inventory elicited less accurate job information as it encouraged job incumbents to engage in gap-filling, and report “extra” items. Although job incumbents were not presented with "original" stimuli, the script task format could have lead them to tick extra tasks in the script that they did not perform, and that they would not have ticked if the tasks were alphabetised by verb. Clearly the results of this study cannot determine which task ordering elicits the most accurate job information, however, the findings of this study highlight a need for future research to examine the accuracy of job information elicited by different task orderings.
Average Relative Time Spent on Tasks and Ease of Task Identification

However, while task order influenced the number of tasks a job incumbent ticked, it did not impact on the average relative time spent on tasks a job incumbent reported. In addition, task order did not influence how easy job incumbents found it to identify tasks in the task inventory that they perform. It may be that task ordering does not impact on how easy job incumbents found it to identify the tasks they perform. Coover (1990) found that when respondents completed task inventories with tasks ordered four different ways, the most preferred order was by script using script headers. However, in the current study job incumbents completed only one task inventory, and so could not be asked which task ordering they preferred. The results obtained here compliment the work of Coover, and are useful as they show that when job incumbents are asked to make a subjective rating on how easy it is to identify the tasks they perform, task order does not affect the response provided.

Practical Utility of Results

Occupational Structure

The findings in this study indicate that the order in which tasks are presented could influence the overall occupational structure identified. As the script group ticked significantly more tasks than the alphabetised group, yet reported a similar average relative time spent on tasks, the overall time spent figures for the script and alphabetised groups would be different. This in turn, would lead to a different clustering solution being obtained for the script and alphabetised groups. Different clustering solutions, however, cause problems as it now becomes questionable as to what the “actual” job groups are. It was beyond the scope of this study to examine the job groupings obtained by both the script and alphabetised inventories. However, the results of this study suggest that task orderings should be kept consistent (that is, use script based or alphabetised by verb), as different sets of job groupings may be obtained when tasks are ordered by script versus alphabetised by verb.

Remuneration Reviews

At a remuneration review, the average number of tasks performed within a trade, or the overall amount of time job incumbents within the trade spend on a specific task might be examined (M. Holmes, personal communication, September 12, 2002). The results of this study, indicate that the task ordering used within a task inventory could influence the outcome of a remuneration review. For example, if it needed to be demonstrated that a trade is very diversified, a script based task inventory would most suitably achieve this outcome. A script based inventory would be most suited as it gets job incumbents to tick more tasks than an inventory that is alphabetised by verb. However, if a pay case was to be based on job incumbents within a trade spending a lot of time on a few tasks (due to a specialisation), an alphabetised by verb task inventory would be the tool of choice. An alphabetised by verb inventory would lead job incumbents to tick fewer tasks than the script based inventory, consequently in the alphabetised inventory a higher overall time spent rating is assigned to each task. Clearly given that task ordering could have an impact on practical outcomes of OA, it is important to determine which task ordering elicits from job incumbents the most accurate information about their job.
Limitations

One of the limitations of this study is that the script-headers used were too broad; hence, scripts were not fully activated for job incumbents completing the script based task inventory. In this study it may have been erroneously assumed that the hierarchy of a script and the hierarchy within a task inventory are analogous. Given that some script effects were evident in this study, it may have been that the potential for any large benefit resulting from ordering tasks by scripts was lost, as the broad duty area headings were not specific enough to act as script-headers. It is therefore suggested that in large task inventories, sub-duty area headings containing only one or two scripts may be required to trigger full script activation within job incumbents.

Other limitations to this research relate to the extent to which these results can be generalised, and the absence of recent literature in this area. Firstly, these results are pertinent to the Australian and United States Defence Forces where the alphabetised by verb task ordering is adopted. However, they cannot be fully generalised to the Canadian Defence Force as the format of the script based inventory used in this study differed from that used by the Canadian Defence Force. Secondly, studies by Coovert (1990) and Getkate (1994) represent the limited amount of research on how cognitive processing can explain how respondents process task information. Furthermore, most of the basic and generic script literature that this study relied on was conducted over 20 years ago. Nevertheless, despite the above limitations, the results from this study provide valuable practical insights for the future of OA.

Conclusions

This research presents original findings on how task ordering can affect the job information provided by job incumbents. This study represents a valuable contribution in the area of OA, particularly as it represents the first study undertaken in an applied setting, which evaluated whether different task orderings affect the job information provided by job incumbents. There were two key results to emerge from this study. Firstly, that job incumbents exposed to a script based task order ticked significantly more tasks than job incumbents who completed a task inventory where tasks were alphabetised by verb. Secondly, senior ORs ticked significantly more tasks than their junior OR colleagues, indicating a different work role, and consequent different work script being adopted by senior and junior ORs. Given that a task order effect was present, the results have significant practical implications, especially in the areas of occupational structure development and remuneration reviews. Furthermore, the presence of an order effect raises the biggest question to be addressed in future research – which task ordering elicits the most accurate job information from job incumbents?

---

23 The Canadian Defence Force does not include duty area headings in the final task inventory. However, duty area headings were used in the script based task inventory in this study, as removal of duty area headings from only the script based task inventory would have caused ‘duty area heading’ to be a confounding variable.
However, in light of the current findings it would seem that when the population to complete the task inventory is large and heterogeneous, a task inventory where tasks are alphabetised by verb is most appropriate. Whilst for smaller homogeneous samples, there may be benefits in using a task inventory where tasks are ordered by scripts using script-headers. Task inventories have been used to collect information about jobs for over 80 years, and perhaps will be used for the next 80 years. However, in an increasingly demanding market, job analysts are challenged to collect, with minimal impact, the most accurate information from job incumbents about their jobs.
References


THE MULTI-SKILLED SOLDIER CONCEPT: CONSIDERATIONS FOR ARMY IMPLEMENTATION

Elizabeth J. Brady, Ph.D.
Senior Research Psychologist
U.S. Army Research Institute for the Behavioral and Social Sciences
brady@ari.army.mil

Introduction and Overview

This paper provides an overview of the Multi-Skilled Soldier (MSS) Concept Study, Phase I, designed to explore the nature of individual soldier skills required from now through the full fielding of the Objective Force. Study objectives included the institutional implications in seeking a more multi-skilled soldier, a generic timeline to implement a MSS Concept by 2008, and salient behavioral research questions that must be addressed as part of an effective implementation process. Each of these objectives will be addressed briefly starting with the issues surrounding the MSS Concept currently under consideration by the Army to include, but not limited to, various perceptions of the MSS, notional definitions of the MSS, and training implications.

The published study report is entitled The Multi-Skilled Soldier Concept: Considerations for Army Implementation (Nelsen and Akman, 2002). This study relied heavily on interviews with those involved with developing the concepts for the Objective Force (OF) and with those connected with fielding the Army’s Interim (now Stryker) Brigade Combat Teams (SBCTs).

The Multi-Skilled Soldier Concept

The MSS Concept is still clearly embryonic. Throughout the Army, different organizations and individuals have different perspectives about it. This section summarizes the various views that have emerged and concludes with an initial characterization of an MSS Concept. This characterization will then serve as the basis for an MSS Developmental Blueprint, MSS Roadmap, and MSS Study Plan.

Traditional Army Approaches to “Multi-Skilling”

The notion of developing soldiers with increasingly broader sets of basic skills is not new to the Army. Unit commanders have frequently attempted to cross-train their
soldiers so that they can expertly handle all of the weaponry in their platoons. In combat, for example, if an assigned machine gunner became a casualty, another soldier could step forward and immediately operate the machinegun proficiently.

Unit commanders have also habitually resorted to assigning soldiers to positions requiring largely different sets of skills than those of their basic Military Occupational Specialty (MOS). In these cases, the soldiers generally learn their new skills through on-the-job training, whether formally or informally, and are usually awarded a secondary MOS upon demonstrating proficiency in the new position. This is usually done principally for the needs of the unit, particularly when shortages in a critically needed MOS occur.

Another category of cross training involves teaching soldiers to perform carefully chosen sub-sets of tasks belonging to an entirely different MOS. For example, many units are required to train and certify a certain number of their soldiers as combat lifesavers. These soldiers are trained in advanced combat lifesaving, first-aid tasks that, in essence, would encompass many of the apprentice skills required of a trained Health Care Specialist assigned to the battalion’s medical platoon. Having some combat lifesavers in the unit means that a seriously hurt soldier will more likely receive immediate care.

The Emerging Concept of the Multi-Skilled Soldier

Below are four possible approaches to consider in broadening soldier skills. These approaches tend to focus more on soldiers during their initial training experience rather than the traditional approach (after initial entry training or before one becomes a Non-Commissioned Officer).

1. The **Additionally Skilled Soldier** approach focuses on training soldiers differently during their initial training base experience [the equivalent of today’s Basic Combat Training (BCT) and Advanced Individual Training (AIT)]. Soldiers would report to their initial unit of assignment cross-trained in a carefully crafted blend of tasks considerably different from those traditionally associated with their primary MOS. This approach would be linked directly to the operational requirements of the future OF.

2. The **Generic Soldier** approach focuses on continuing MOS consolidations. Soldiers would become more MOS generic in the sense that they would reflect a selective consolidation of major MOS along the lines recently implemented for the infantry MOS to form a single generic infantry MOS, the 11B (consolidation of the 11B traditional Infantryman, the 11M Fighting Vehicle Infantryman, and the 11H Heavy Anti-Armor Weapons Infantryman). Using this methodology, the same kind of consolidation may be used to create other, more generic MOS, significantly reducing the number of MOS.

3. The **Adaptable Soldier** approach focuses on developing soldiers with significantly enhanced capacities for adaptability, versatility, and mental flexibility. The notion is that a soldier’s training provides not only skill sets to perform tasks effectively in given contexts, but the necessary attributes and orientation to adaptively, and perhaps creatively, apply existing knowledge and skills to deal successfully with new tasks in different situations. For many, this approach is termed “competency-based,” implying
the ability to “get things done,” whatever they are, by adaptively employing existing inventories of knowledge, abilities, attitudes, and skills.

4. The Perpetually-Learning Soldier approach is referred to as that of “life-long learning.” Soldiers must be educated, trained, and motivated to broaden the skills they acquire in their initial training experience and thereafter through self-education [e.g., distance learning (DL)].

This Study’s Characterization of the Multi-Skilled Soldier

The impetus for the MSS is to have a soldier who is extremely adaptable and can perform a greater variety of tasks on the battlefield. While having soldiers with more skills will facilitate that, the soldier may also need to have more abilities and more knowledge. The key is establishing the requirements for what soldiers must be able to do in future Units of Action. Defining those requirements will, in turn, lead to an identification of the mix of Knowledge, Skills, and Abilities (KSAs) that the soldier must have.

Understanding tasks and KSAs help explain the various depictions of the MSS. This soldier may have more KSAs than presently. Ultimately the goal is to have soldiers who can perform more tasks. In essence, the MSS, if the U.S. Army’s soldier goals are to be achieved, is a multi-KSA soldier. This study characterizes the MSS Concept as follows:

1. First and foremost, the MSS is one who may be recruited from a pool having a different ability set than current recruits and who emerges from the initial training experience better trained in the base MOS knowledge and skills required for his first troop assignment than today’s soldier. The training would be more generic, as discussed above, resulting from a degree of MOS consolidation from today’s MOS structure.

2. The “generic” training that soldiers receive for their base MOS would emphasize a “competency-based” training approach wherever possible. The soldier will train on representative equipment systems, with a primary emphasis on how to adapt rapidly to similar but different systems throughout his functional area. The intent is to produce a more “thinking soldier,” partially by employing different selection criteria and partially by implementing a mainstream training approach which is much less “rote-based” and much more “competency-based” or “comprehension-based.”

3. At the same time, soldiers would be trained on additional skills and taught additional knowledge that fall outside the scope of today’s base MOS. The soldier would receive carefully selected subsets of apprentice-level skills and knowledge associated with other MOS, most of which are associated with other branches in today’s frame of reference. The main purpose is to provide increased combined-arms skill depth and redundancy, particularly for low-density MOS, in Table of Organization and Equipment (TOE) units. The principal driving force for the groupings of these skill sets would be the operational needs of the OF’s Unit of Action. However, the MOS design would also promise to enhance combined-arms skill depth and redundancy for Interim and Legacy Force units as well.

4. Once he joins his first TOE unit, the MSS would be expected to sustain and deepen his skills and knowledge, relative to both his base MOS and the additional sets of skills and knowledge he received from initial entry training. Much of the effort to do so
would fall on the shoulders of the individual soldier, who would be expected to pursue skill and knowledge growth both through unit training experiences and energetic self-study, validating “success” via tests and other certifications. Career development incentives would reward soldiers who do so and penalize those who do not.
Potential Impacts on the Training Paradigm

One of the areas most affected by the implementation of the MSS Concept will be the training base, particularly that part of the Institutional Army responsible for conducting initial entry training and all subsequent “schoolhouse-training.” In the future, the Institutional Army’s role in distributed training will become increasingly pivotal for sustaining and broadening soldier skills after initial entry training. There are many options for the potential implementation of the MSS Concept in the Army. For example, the MSS Concept could be implemented only for OF units. Or, it could be implemented for all soldiers undergoing initial entry, regardless of whether or not their follow-on assignments will be to OF, Interim Force, or Legacy Force units. In addition, the MSS Concept could be implemented over a short period of time in the next few years, or it could be implemented in incremental, evolutionary fashion in phases over time. In short, different options will have different impacts on the training base.

MSS Developmental Blueprint

The study describes an MSS Developmental Blueprint, which frames a process that can be used to define the salient issues and factors the Army will have to address in formulating and assessing options, or courses of action, for implementing and sustaining an MSS Concept. The purpose of the MSS Developmental Blueprint is to provide Army personnel and training planners a process model to identify the key elements, tradeoffs, and constraints necessary to make decisions leading to multi-skilled soldiers. Five major components are featured in the MSS Developmental Blueprint: force design, MOS design, training design, implementation, and sustainment. This process starts with a determination of the force design and proceeds to crafting an MOS structure to support the requirements of that force design. The results drive an assessment of the implications for the existing training paradigm and the adjustments, or even re-design, that may be required to implement the new MOS design. The next components address the considerations for implementing and sustaining the new MOS design throughout the institution.

MSS Roadmap

An MSS Roadmap indicates the steps required in the short-, near-, and long-term to emplace a functioning MSS system by 2008, at which time the first units of the Objective Force should enter the force structure. Activities fall within the purview of each of the three major organizational groupings that would play a leading role in the process, namely Army Leadership and the Personnel and Training Communities. The timeline also reflects the existence of the three different Army forces (Legacy, Interim, and Objective) that will co-exist for many years after MSSs start entering the force. Changes to the Army’s job structure must occur in a way that supports the operational requirements of all three of these forces, as well as of the Army Reserve and Army National Guard.
MSS Study Plan

One of the challenges facing the Army is that MSS implementation requires background information from research that has not yet been conducted for the MSS Concept. An MSS Study Plan identifies research issues that, when addressed, will contribute significantly to the development and implementation of the MSS Concept. Research projects (basic and applied) were organized based on requirements potentially facing Army Leadership and the Personnel and Training Communities.

Conclusions and Recommendations

First and foremost, there is no official consensus or definition of the MSS relating to the requirements of the OF. There is a general consensus that OF soldiers should be more adaptable and provide more skill depth and redundancy than present-day soldiers. However, there is no authoritative working definition of the multi-skilled soldier, no layout of MSS expectations in OF requirements documents, no integrated implementation plan in place or under development, and no proponent clearly charged with developing one. The Army needs an authoritative definition of the MSS Concept.

Secondly, there is no integrated planning underway for comprehensive MSS implementation. There is no Army-wide proponent for MSS Concept development. Implementing the MSS concept will require a step-by-step process in which concepts and plans are more clearly adjusted and refined at each successive stage. While there is a broad consensus that the MSS Concept is an essential ingredient in the human dimension design of the OF, most of the efforts associated with the OF have thus far focused on operational concepts, materiel acquisition, and organizational design. It is recommended that the Army establish an overall proponency for MSS realization, select a corresponding course of action, and put an integrated master implementation and sustainment plan in place.

Third, most of the discussions about the MSS Concept have focused on the needs of the OF. But, it is important to remember that what are today termed the OF, the Interim Force, and the Legacy Force will co-exist for many years. The benefits of MSS implementation can be applied to increase skill depth and redundancy in Interim and Legacy Forces as well. Designing and implementing the MSS Concept for the entire Army, including the Reserve Components, is much more efficient and desirable than operating different training and personnel systems for the Interim and Legacy Forces. It is therefore recommended that the Army should structure MSS implementation and sustainment for optimal and parallel benefit of the Interim and Legacy Forces. This means designing an MSS implementation and sustainment approach from the beginning that, insofar as possible, can be applied across the entire force, to include the Reserve Components.

Finally, the MSS Concept involves issues of human cognitive potential and associated practices that have not been adequately researched for the requirements of the OF. There are a number of unanswered questions that will impact virtually every aspect of the life cycle of the soldier to include: recruiting, classification, selection, distribution,
deployment, retention, training and education, sustainment, professional development, and separation. Timely behavioral research can help ensure that the initial MSS designs are within the realm of sound cognitive expectations for the general soldier population. It is therefore recommended that the Army conduct the necessary supporting behavioral research to implement the MSS Concept effectively and efficiently. Without such fundamental research, pursuing MSS could easily become a high-risk, “trial-and-error, hit or miss” venture.

**The Multi-Skilled Soldier Concept, Phase II**

TRADOC’s PPD requested further development of concepts leading to the eventual implementation of MSS as part of the Interim and Objective Forces. While the just concluded study provided the first formal effort to define the issues and concepts, there is a need to refine them, and to investigate in greater depth concept applicability to the Interim and Objective Forces.

The Multi-Skilled Soldier Concept, Phase II study began in April 2002, with a projected end date of January 2003. The objective of this contract is to develop a more definitive, operational MSS concept that can serve as the basis for development and implementation efforts. The first published report of this effort is entitled *Principles for Defining Multi-Skilled Jobs Based on Mission Requirements of Multi-Functional Units: The Multi-Skilled Soldier Job Modeler (MJM)* (Akman and Nelsen, 2002).

**References**


Career Transitions Within Members of the Canadian Armed Forces

Captain Karen Brown
DHRRE

In today's fast paced world, career change is prevalent. Old standards of maintaining a single life-long career are often falling by the wayside as individuals search for personal satisfaction and meaningful work. Career transition, the process of changing careers and the adjustment to a new career, can be complex, stressful, and difficult to manage. As a result, career transition research has recognized the need to understand psychological resources and barriers that adults possess during a career transition (Heppner, 1998).

Within the Canadian Forces, there is a continuous flow of members in the process of career transition due to the fulfilment of contract, medical releases, voluntary releases, and, in the past, through the Force Reduction Program (FRP).

In order to increase retention, there is a high interest in understanding releasing members’ decisions to leave. Despite concerns about high levels of attrition, the Canadian Forces remains dedicated to ensuring that released personnel are prepared to face the challenges involved in a career transition into the civilian sector. Accordingly, the goals of this study were to determine: (a) why members of the Canadian Forces chose to leave; (b) what, if any, was their level of depression during this transition and what effect did this have on their career transition; and (c) what psychological resources (i.e., control, readiness, confidence, decision independence, and support) did the participants perceive during this transitions.

Life Factors. Life factors, which play a big role in the decision to change or maintain a career, include: (a) job satisfaction, (b) work environment, (c) personal relationships, (d) perceived support, (e) financial situation, (f) family responsibilities; and (g) perception of
alternate career (Cherniss, 1989; Neapolitan, 1980; Rhodes & Doering, 1993; Heppner, Multon, & Johnston, 1994; Lenz & Reardon, 1990). This list of predetermining factors provides an understanding of factors affecting the decision to change careers, but is no means exhaustive. In addition, certain demographic characteristics have predicted career transition (e.g., gender, age, and marital status; Wheaton, 1990). One method to increase retention is to determine the major factors that contribute to decisions to leave the military. As such,

*Research Question 1:* What do former members of the CF consider the most important factors in their decision to change careers?

*Research Question 2:* Do demographic characteristics predict reasons for release?

**Career Transition and Depression.** Mental health has been found to influence career transitions. For example, Wheaton (1990) found that personal history, prior to the transition (e.g., prior work problems, gender, age, and marital status), influenced the stressfulness of the event. Heppner, Cook, Strozier, and Heppner (1991) addressed the effects of career-related variables (e.g., vocational identity and career barriers) and coping variables (e.g., problem-solving appraisal) on depressive symptomatology and perceived stress in the career transition of farmers. For female farmers, depressive symptoms were best predicted by perceived barriers to career change (Heppner et al., 1991). For male farmers, an emotional coping style was most predictive of depressive symptoms. Also, a lack of confidence in the ability to solve problems predicted higher levels of stress, reduced progress in career transition, and increased depression. While this study underlines the need for further research in gender differentiation in career
transition coping styles and depression level, it also demonstrates the importance of considering the mental state of individuals during career transitions. Accordingly,

Research Question 3: What is the present level of depression in former members of the CF?

Career Transition. Schlossberg’s model (1984) of transition considers the influence of external factors (e.g., age, race, and sex) and internal factors on individuals’ ability to a life transition such as career transition. He purported that individual differences facilitate one’s reaction to life transitions. In order to assess and bolster internal factors, Heppner et al. (1994) developed the Career Transitions Inventory (CTI) to measure the perceived psychological resources at the time of a career transition. The goals of this 40-item, multidimensional inventory are: (a) to help individuals understand the psychological factors that influence their career transition, (b) to provide a common language for counsellors and clients to discuss the career transition, and (c) to assist counsellors in the development of career counselling strategies that are targeted for the individual’s need (Heppner, 1998). The five factors that compose the inventory are: (a) readiness, appraised motivation for making a career change; (b) confidence, felt towards successful career transition; (c) control, the degree the individual feels the transition is under their control; (d) support, social support for the transition; and (e) decision independence, level of autonomy to make this decision. The CTI has practical applications for career counsellors and can assist in determining the perceived psychological resources and barriers present. Thus,

Research Question 4: What are the levels of psychological resources for military members during the career transition?
Hypothesis 1: The level of depression will be negatively related to levels of psychological resources.

Method

Fifty former members of the CF voluntarily participated in the study by completing the Career Transition Survey. The majority of the subjects were male (76%), non-commission officers (92%), and between the ages of 36 to 43 (40%). Forty percent of participants had served more than 21 years with the CF and 36% many had served 16-20 years. The participants were distributed over 22 occupations. The Career Transition Survey was composed of 100 questions: (a) Reasons for Leaving the Military, 25 items; (b) Service Background, 14 items; (c) Beck's Depression Inventory, 21 items (Beck et al., 1961); (d) Career Transition Inventory, 40 items (CTI; Heppner et al., 1994); and (e) Mood De-induction Task (Life Events Inventory; Sinclair, Mark, Enzle, & Borkovec, 1994). Average completion time was 30 minutes.

Results and Discussion

The most important factors in the participants’ decisions to change careers were determined by frequency levels of 25 participants or greater. The following were considered the most important factors that affected the decision to leave the CF for participants: (1) Force Reduction Program, (2) For greater achievements/challenges, (3) Dissatisfaction with military life, (4) A need for a greater sense of personal fulfilment with work, (5) To spend more time with family, (6) Lack of promotions, (7) A desire to pursue a new trade/occupation, and (8) To have a more challenging job and/or career.

The importance of factors affecting the decision to leave the military was further analyzed by a repeated measures analysis of variance with reasons (MONEY x
INTRINSIC x NEITHER) as the within subjects factors. There was a main effect of reasons, $F(2, 48) = 23.60, p < .001$. Intrinsic factors ($M = 3.34, S = 0.98$) were rated significantly lower than monetary factors ($M = 3.99, S = 0.57$), $Tukey HSD = .286, p < .001$. These results reveal that monetary reasons had a greater affect on the decision to leave the military; this contradicts findings where intrinsic motivation was rated higher (Cherniss, 1989; Kanchier & Unruh, 1989).

The relationship between demographic characteristics and reasons to leave the CF were analysed with independent t-tests with a Bonferroni cut-off value of $p < .004$ used to correct for the possibility of Type I errors. Interesting results included: (a) females rated the number of operational tours, a desire to reduce stress, a desire to increase family time, and undesirable working conditions higher than males; (b) members with less than 20 years rated labour market, a desire for a challenging job, a need for greater achievement, and personal fulfilment, higher than those with more time in the CF; and (c) members who had gone on operational tours placed greater importance on having a civilian job waiting for them. Demographic characteristics had no relationship with level of perceived psychological resources or levels of depression.

The majority (90%) of participants scored in the normal range for level of depression, while a mere 4% were considered mild to moderately depressed and only 6% scored in the level of moderate to severe depression.

The majority of participants (66%) obtained an average score (i.e., 141 - 198) for their level of psychological resources, and the remaining participants (34%) scored above average (199 – 228) and no participant had a below average (i.e., 112 - 140) level. The mean of the overall CTI score for military members was higher ($M = 188.91, SD = 17.79$).
than that originally found by Heppner et al. (1994; $M = 170.66, SD = 28.56$), $z = 4.52, p < .05$. The difference in levels may be attributed to the dissimilarity between the two populations. The most dramatic difference was the higher level of perceived confidence for this sample ($M = 54.10, SD = 6.49$) versus ($M = 43.56, SD = 8.21$), $z = 2.93, p < .05$. This indicated that the former members of the military were confident they would succeed in their career transitions. Military members scored significantly higher on all levels of psychological resources with the exception of perceived level of decision independence, which, although higher than Heppner et al.’s (1994) population, was not significantly higher.

A significant negative correlation between depression level and psychological resource level (i.e., overall CTI scores) was found ($r^2 = -.29, p < .05$). However, these results should be interpreted with caution due to the homogenous nature of levels of depression. No significant relationships were found between depression levels and individual factors of psychological resources (e.g., readiness). Thus, previous results (e.g., Heppner et al., 1991; Wheaton, 1990) cannot be fully supported due to the low levels of depression, which indicates that the majority of participants suffered no depression at the time of completing the survey. Conversely, high levels of perceived psychological support may have lowered any potential depression normally experienced during career change; thereby, supporting Heppner et al. (1994) contention that psychological resources aid the individual during a career transition.

This study identified the top reasons that participants chose to leave the military, which varied significantly across different demographic characteristics (e.g., gender and age). The low level of depression indicates positive mental state of the participants that could
be due to psychological resources or the support received from the military in preparing them for a second career (e.g., career counselling, workshops for interview skills and job search techniques, and financial and resume writing benefits). Conversely, the negative relationship between depression and psychological resources provides support for the practical utility the CTI for career counsellors. Finally, the need to understanding the reasons leading to and the factors influencing the outcomes of career transition are important to organizations (to understand attrition), to career counsellors to provide tailored programs, and to individuals facing career transition.

References


ABSTRACT

With respect to the demands of our (post-) modern world one can observe an increasing number of problems among many young people now classified as adjustment disorders. This becomes obvious also for the military service in the Austrian Armed Forces. The clinical psychologists of the Psychology Service are increasingly confronted with this phenomenon. Since the manpower of the Austrian Armed Forces is still based on compulsory national service, all Austrian male citizens reaching the age of nineteen are obliged to service, whether they want to or not. The only choice they have is between military service (8 months) and civilian service (11 months). Therefore a lot of young men get into military service, who would never have chosen this voluntarily. After describing the typical life situation of young Austrian men (average age 19 years) a psychological model to understand life change crisis situations will be presented. Focusing on the criteria of adjustment disorders according to the Diagnostic and Statistical Manual of Mental Disorders (DMS IV), a comprehensive conception including empirical data obtained from the Austrian Army will be proffered. Finally, possible intervention strategies to help young recruits to cope with adjustment disorders will be discussed.

INTRODUCTION

In the year 1995 the Psychology Service of the Austrian Armed Forces established a ‘Help Line Service’ for all people working and serving in the army. This service is thought for emergency situations when people feel out of personal control to handle their life and do not know what to do. The help line service is based on a team of about 30 professionals (military psychologists and other special trained people of the army). The helper who has to do his or her service is equipped with a mobile telephone and can be contacted at any time (24 hours service). Most of the callings are stemming from the young recruits who show a wide variety of problems. These problems are ranging from more or less simple ones (e.g., being not able to spend the weekend at home) till the announcement of a suicide attempt. The last one fortunately only represents a minority.

The team members of the help line service are more and more confronted with a new problem situation. Young recruits (mean age 19 years) are contacting the help line because they are not able to cope with the new situation in the military field. They tell the helper that they cannot cope with the military service, that they are anxious, despaired and demoralized. They wish to leave the army as soon as possible. Even the mothers of
the recruits contact the help line telling about the problems of her sons (e.g., “My son is sitting in his room and crying. He cannot bear the military service”). The helper is confronted with a wish of the recruits he or she cannot fulfill. It does not belong to the policy of the Austrian Army to drop out all recruits suffering on their changed life situation. But at the same time the helper has to search for possibilities how to help the recruits to master their problems. Before going more in detail some empirical data are offered to give some impressions about the frequency of the appearance of adjustment disorders.

Between 1995 and 1997 a total number of 992 recruits were referred to the Ambulance for Clinical Psychology of the Military Hospital at GRAZ. 328 of them (33.1%) were diagnosed by DSM-IV criteria as suffering from Adjustment Disorder (AD) due to non-combat military stress. The percentage of occurrence of AD in this sample is slightly higher than the average seen in psychiatric consultation services for adults (5%-21%; Despland, Monod & Ferrero, 1995). How can this phenomenon be understood and explained?

The literature about adjustment disorders is very rare (e.g., Horowitz, 1997) in comparison to other clinical disorders. For getting a first survey of the complex phenomenon of the adjustment disorders the DSM IV (APA, 1994) criteria are used.

DESCRIPTION AND DEFINITION OF ADJUSTMENT DISORDERS

Symptoms

The development of emotional or behavioral symptoms in response to an identifiable stressor(s) occurring within 3 months of the onset of the stressor(s).

These symptoms or behaviors are clinically significant as evidenced by either of the following:

*) marked distress that is in excess of what would be expected from exposure to the stressor
*) significant impairment in social, occupational for than educational functioning

The stress-related disturbance does not meet the criteria for another specific disorder. Once the stressor (or its consequences) has terminated, the symptoms do not persist do not persist for more than additional 6 months.

Adjustment Disorders Subtypes

With Depressed Mood
With Anxiety
With Mixed Anxiety and Depressed Mood
With Disturbance of Conduct
With Mixed Disturbance of Emotions and Contact
Unspecified

**Associated Features**

Depressed Mood
Somatic/Sexual Dysfunction
Guilt/Obsession

**Differential Diagnosis**

Some disorders display similar or sometimes even the same symptom. The clinician, therefore, in his diagnostic attempt has to differentiate against the following disorders which one needs to be ruled out to establish a precise diagnosis.

**Personality Disorder**
Not Otherwise Specified Disorders (e.g., Anxiety Disorder Not Otherwise Specified);
Posttraumatic Stress Disorder, and Acute Stress Disorder
Psychological Factors Affecting Medical Conditions; Bereavement;
Non-pathological Reactions to Stress

**Cause**

Many people have difficulties adjusting to stressful events. Examples of stressful events include starting a new job, ending an important relationship, or conflicts with work colleagues. As a result, the individual may have difficulty with his or her mood and behavior several months after the event. Some people who have recently experienced a stressor may be more sad or irritable than usual and feeling somewhat hopeless. Other individuals become more nervous and worried. Still other individuals combine these two emotional patterns. Sometimes individuals become reckless and have difficulties following societal rules. The symptoms associated with adjustment difficulties usually subside about 6 months after the stressful event. (Description and definition summarized from: American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders, fourth edition.* Washington, DC: American Psychiatric Association.)

The DSM IV provides a description, classification, and explanation of adjustment disorders, but offers very little for a deeper understanding of the underlying generating principles and for planning of psychological treatment.

**APPROACHES TO UNDERSTAND ADJUSTMENT DISORDERS**

There are two basic but different approaches to understand adjustment disorders. The first one may be called the atomistic (traditional) approach; the second one may be called the systems approach. The basic assumption of the atomistic approach states that an individual may develop a more or less circumscribed stable adjustment disorder in
response to an identifiable stressor. The disorder is considered as a kind of an entity located within the individual. This approach represents the DSM IV conception. In contrast, the basic assumption of the systems approach states that individuals create or remove adjustment disorders with respect to situation-specific demands and subjective needs. The disorder is considered as a relational phenomenon based on a context-sensitive inability of an individual to organize his or her own behavior.

To get a deeper understanding of the systems approach some considerations with respect to human development and coping with stress are presented. These considerations refer to the existential conditions of every day life, to the processes of differentiation and integration embedded in individual-environment-interaction, to personal growth and self-complexity, to some conditions affecting self-derogation, and to main sources of interpersonal problems.

**Existential Conditions of Every Day Life**

All human beings are confronted continuously with two basic dimensions of every day life. On the one hand a lot of demands have to be fulfilled. These demands may stem from outside (e.g., to stand up in the morning, to bring the children to the school, to write a report for the chief) or may be self-defined (e.g., to stop with smoking, to study a language, to learn to play an instrument). On the other hand everyone has to take care for his or her well-being and for fulfilling his or her wishes, desires and interests. Normally, most people don’t have problems to coordinate these two dimensions. They have learned to cope with a wide variety of demanding situations. How will this ability to cope with the many different demands be reached?

**Differentiation and Integration**

The general coping capacity of individuals can be understood as a never ending consequence of two opposed processes: The process of differentiation and the process of integration. According to Watkins (1978) differentiation is the perceptual recognition of differences and discrimination between two different things; integration is the bringing together of two or more elements so that they are consonant or can interact constructively with each other.

**Individual-Environment-Interaction, Personal Growth and Self-Complexity**

The processes of differentiation and integration are embedded in individual-environment-interactions promoting mental (cognitive) development (e.g., various intellectual abilities, value orientation, world view, belief systems, self-reflection), affective development (e.g., affect-regulation, affect maturity, empathy, need satisfaction), and motor development (e.g., various motor skills, sports, dancing). Under optimal conditions these processes will produce personal growth and self-knowledge. According to Linville (1987) self-knowledge is represented in terms of multiple self-aspects. Greater self-complexity involves representing the self in terms of a greater number of cognitive self-aspects and maintaining greater distinctions among self-aspects.
Some Conditions Affecting Self-Derogation

Unfortunately, many people don’t have good conditions for their development and their personal growth. Many years ago Angyal (1951) summarized some basic conditions affecting self-derogation. These conditions also can be found today and are still valid.

A. The over-protective attitude of an insecure, anxious parent tends to convey to the child a feeling that he lives in a world that is full of dangers, and with which he is inadequate to cope. When a parent does too much for the child, he is telling him by implication that he is incapable of doing things by himself.

B. When the parent is too eager for the child to do well and is excessively critical to him, he is likely to instill in the child the feeling “something must be very wrong with me; I can’t do anything right.”

C. When parents distort and exaggerate the child’s achievement, when they cast him into a superior role and have great expectations of him, they plant the seed of self-derogation in still another way. Deep down the child knows that the parents’ picture of him is untrue, and measuring himself by these excessive and often fantastic standards, he secretly begins to feel utterly worthless.

D. The too many “don’ts” which the child hears tend to create in him the feeling that those things which he most wants are the things that are forbidden and evil. This easily can give rise in him to a secret conviction that he is a fundamentally evil person.

E. The ways in which children are being treated without understanding and without respect are many, and these are likely to create in the child the feeling that he just doesn’t matter in this adult world, that he is of no account, that he is worthless. Often one wonders why the child accepts the verdict that he is worthless, instead of blaming the parent for being so obviously lacking in understanding, so wrong and selfish. The answer suggests itself that the child needs so much to feel that he has “good parents” that he tenaciously adheres to this belief and would rather assume himself to be evil or worthless than give up the idea that he has good parents. (p. 138).

Transferring these conditions to the area of adjustment disorders one can assume that some of them have causal influence. Explorations done by the military psychologists show that especially an over-protective attitude of the parents seem to play an important role. There is a further interesting feature. The recruits have grown up under very liberal conditions, even without any rules. This leads to the paradoxical situation, that on one hand the recruits could do what they wanted, on the other hand that they lived in an over-protective clime. As a consequence, they have developed the tendency to avoid uncomfortable situations in everyday life, thereby excluding the important experience how to master stressful events in different areas.
Main Sources of Interpersonal Problems

In clinical practice three main sources of interpersonal problems can be observed: Deficits in social skills, deficits in regulating emotional arousal in interpersonal interactions, and deficits in regulating closeness and distance. Most recruits show a mixed picture of these deficits. For example, some of them are unable to generate self-assertive behavior; others are unable to find contact with their new comrades. It is important to stress again that these deficits are extremely situation-specific.

MANAGEMENT OF ADJUSTMENT DISORDERS

From a systems point of view two basic strategies to treat adjustment disorders have to be distinguished: Individual-centered strategies and environment-centered strategies. The individual strategies include supportive measures (e.g., the recruit gets personal help by the psychologist or by his commander), expressive measures (e.g., the recruit can speak freely about his emotions and his demoralization), exercise, training, and practicing procedures (e.g., learning a relaxation technique), and confrontational techniques. The last ones are done only exceptionally in the psychological ambulances of the military hospitals, because they are normally part of a long-term psychotherapy that cannot be offered by the military psychologists.

The environment-centered strategies include activating external resources (e.g., to find a person outside the military taking care for the recruit), changing situation-specific demands (e.g., the recruit will get another function in his unit), and providing an environment that avoids overload.

CONCLUSION

Adjustment disorders cannot be understood and diagnosed sufficiently within the frame of the atomistic (traditional) approach. Moreover, a systems approach focusing on deficits of the ability to self-organize one’s own behavior which is sensitive to individual differences and to situation-specific demands provides a better basis for the diagnosis and psychological management (treatment) of adjustment disorders.

Thus, effective management of adjustment disorders has to focus on the individual-environment interaction and to provide treatment techniques to the individual which may allow to overcome his or her demoralization (e.g., supportive and expressive measures), to improve his or her self-organizational abilities (e.g., exercise, training, practicing procedures, and confrontational processes) and to activate external resources to reduce environmental stress (e.g., change of the military unit, getting a new function).
REFERENCES


EVALUATING COMMUNICATION IN A SYNTHETIC TASK ENVIRONMENT

2Lt Christopher Hetrick, 2Lt Gary Cooper, 2Lt Kevin Walsh, 2Lt Ryan Smith, and 2Lt Adam Wickes
United States Air Force Academy, Colorado, USA
cdhetrick@hotmail.com

Maj Terence Andre
Air Force Research Laboratory, Mesa, AZ, USA
terence.andre@williams.af.mil

INTRODUCTION

The ability to efficiently and accurately collect and evaluate group communication during task performance is indispensable when analyzing the overall effectiveness of an organization or team. The possibility of automating, or partially automating, the process of collecting and evaluating communication between team members involved in decision-making and problem solving in a synthetic task environment could be valuable in future communications research and analysis. Such a system could possibly be utilized to gain a better understanding of different ways to improve overall team and organizational performance in many different areas of research.

In his article, *Pilot Speak*, Spinetta (2001) describes common properties of effective organizational communication that focus on communication between Air Force pilots, but that he claims also apply to any situation. The first element is that communications should be directive and descriptive. The speaker should tell the receiver what he wants done and how he wants it done. In addition, effective communication should specifically identify who should accomplish the actions contained in the instructions in order to avoid confusion. Transmissions should also be concise and to the point. The key to effective communication is to relay the most information in the fewest words possible (Spinetta, 2001). This idea is reinforced by a study of 237 undergraduate students’ performance in a team tank simulator, conducted by Marks, Zaccaro, and Mathieu (2000), in which they determined that the quality, not the quantity, of team communication is positively associated with team performance. Another important quality of team communication is the support of open communication among team members (Spinetta, 2001). All relevant information is important, regardless of the position or rank of the source. This aspect of communication is a significant consideration, as demonstrated by Palmer and Lack (1995) in a study of crew resource management in Air Force aircraft, which showed that in typical aircraft crews and formations, communication tends to be dominated by the authority figures. Also, communication that is intended to keep the team coordinated and together, such as information relating to progress and location, is important (Spinetta 2001).

COMMUNICATION METRICS

In order to effectively automate communication analyses, communication metrics must be established. A study conducted by Dutoit and Bruegge (1998) quantitatively measured communication traffic on electronic bulletin boards in a problem solving environment by recording the number of messages sent by each team, the number of noun phrases contained in each message, and the number of unique noun phrases. The data was then analyzed using a set of natural language processing tools. The study showed that good communication metrics evaluate the information flow by measuring the volume and complexity of information exchange. The
Common factors found by Dutoit and Bruegge (1998) in their experiment were word counts, transmission counts, noun counts, and unique term counts.

A second area of focus in developing an automated system is communications recording. Oviatt (2000) described the primary concerns regarding automated speech-recognition programs. The first aspect she discussed was the program’s ability to effectively process natural language. A study by Furman and Cosky (1999), explained that the most accurate level of software utilizes grammar-based speech recognition. Grammar-based programs utilize principles of grammar to deconstruct spoken words and reconstruct them for processing, which allows programs to recognize and record natural language and speech patterns.

Another important aspect of an effective speech-recognition program, especially in analyzing communication, is the program’s ability to process dialogue (Oviatt 2000). One of the chief challenges in this area is the program’s ability to discern between the individuals engaged in the dialogue in order to accurately record who is speaking. A significant consideration is the program’s ability to handle speakers stepping on, or interrupting, each other (Furman & Cosky, 1999). Closely related to the program’s ability to handle dialogue is the program’s capability in regards to multiperson use. One of the chief challenges with speech-recognition programs deals with recognizing different individual voices, and at this point in time, most have to be trained to recognize voices that are intended to be recorded (Oviatt, 2000). The final aspect of voice-recognition software deals with error handling. It’s important to realize that voice-recognition software is not perfect and users must devise a method to identify and sufficiently deal with recording errors (Oviatt, 2000).

This research primarily focused on developing communications metrics and incorporating with a voice recognition program. In selecting an effective voice recognition software system, we had to be aware of the potential risks inherent in the systems. We knew that we had to focus on such aspects as processing power, grammar-based recognition, individualization ability, and error handling when selecting our system. One of the greatest challenges of the project was developing communication metrics, primarily because it is a fairly new and underdeveloped area of research that is highly subject to the specific situation to which it’s intended to be applied.

METHOD

Participants
The population sample consisted of eight, two-person teams composed of members from all four classes at the United States Air Force Academy who volunteered to participate in the study. Each team was randomly selected with subjects varying in experience in the performance of synthetic tasks.

Materials
The system consisted of Dragon Naturally Speaking voice recognition software, a synthetic task environment for the subjects to complete, and a component to measure communication metrics. The synthetic task environment selected was Commandos 2: Men of Courage, due to its high degree of user-friendliness and the relatively low amount of gaming skill required to complete the synthetic task. In addition, the division of tasks among the users during game play forced team members to work together to successfully complete the mission, which ensured adequate verbal transmissions to evaluate team communication. Dragon Naturally Speaking was selected because it was the best available option at translating verbal communication into a written document. However, as a precautionary measure, a tape recorder was also used to make backup recordings of team communications.

The communications metrics was developed based on the aspects of quality communication described by Spinetta (2001) and aspects of communication measurement described by Dutoit and Bruegge (1998). Dutoit and Bruegge (1998) explain that the most simple, yet effective, measures of communication involve word counts, transmission counts, noun counts, and unique
term counts, most of which are possible to achieve through automation. Qualities of effective communication described by Spinetta (2001) are: directive, descriptive, and concise transmissions, clear identification of who should carry out the operation, open communication between team members, and communication intended to establish progress. To combine these two approaches for evaluating communication, we modified a word and transmission count method in order to apply it to the qualities described by Spinetta (2001).

First, we performed simple word and transmission counts for each subject and each group as a whole. We then counted verbs in order to assess descriptive aspects of the communication and directive orders (i.e. “Search him and get the cigarettes”) to assess the directive aspect of the communication. Next, we counted identifiers (i.e. “You,” “I,” or names) to assess clear identification of who was to carry out various operations. We also measured statements and questions intended to measure progress (i.e. “Where should we go now?”). Finally, we divided the word count by the transmission count for each subject to assess conciseness in the communication and found the word count ratio between subjects to measure the balance of communication between subjects.

Procedure
Testing was conducted in the US Air Force Academy library computer lab using two Pentium III computers with LAN network connections for the multiplayer synthetic task and two Pentium III laptop computers for the voice recognition software. The subjects began the experiment by training their voices to the voice recognition software for approximately ten minutes. The next step involved a demonstration of the controls and actions that were required to complete the synthetic task. For this training, a demonstrator played through the synthetic task, explaining what he was doing, why he was doing it, and how he was operating the characters while the subjects watched. Following the voice recognition and synthetic task training, the subjects were set up on the system and allowed to start the mission.

The mission consisted of two playable characters, each with unique capabilities, all of which were required to successfully complete the level. Each character was solely assigned to one of the two subjects. Each group was allowed to play until they completed the mission, with restarts allowed using the quick-save function on the game. Their communication was recorded using the voice recognition software and by holding the tape recorder in between the two group members. During game play, researchers evaluated each group on their performance using six measures: Survival, completion time, secrecy, non-lethal tactics, completed objectives, and completed secondary objectives. At the completion of the task, each subject’s voice recognition software recording document was saved to disk.

Since the communications transcriptions provided by Dragon Naturally Speaking were so inaccurate that they were ultimately useless, the data was prepared by transcribing each group’s communications from the tape recording to a word document. The transcription process took approximately 90 minutes to complete per group. Once the communications were transcribed, each transcription was evaluated using the word count function in Microsoft word and by marking and physically counting the number of transcriptions, nouns, directive orders, progress statements/questions, identifiers, and by calculating the conciseness and balance of communication. This aspect of the data preparation process took approximately 60 minutes per group.

RESULTS
A preliminary data analysis was performed using Microsoft Excel. To evaluate the usefulness of each aspect of our communications metrics, we ran a correlation test between each group’s communication measurement and their overall score on the game. Initially, we found weak, but significant, correlations for identifiers ($r^2 = .30$) and conciseness ($r^2 = .35$). However, upon closer inspection of the data, it appeared that two outliers were significantly affecting our data. One was the best group that took nearly half the time of the other groups and the other was
the worst group, which took significantly longer. In such instances, the time alone required for each group significantly affected the word count, thus making it no more valid than simply timing the game as a predictor of performance. After removing these two outliers, we found significant correlations for word count ($r^2 = .59$), transmission count ($r^2 = .54$), noun count ($r^2 = .68$), and directive statements ($r^2 = .39$). These results are shown in Figures 1 through 4, respectively.

![Figure 1: Performance correlation based on word count and score (outlier removed).](image1)

![Figure 2: Performance correlation based on transmission count and score (outlier removed).](image2)
The three primary objectives of this project were to develop communication metrics, evaluate the potential of voice recognition software, and to decide if using the two together to analyze team communication would be a viable option to explore for communication metrics. Following initial evaluations, some aspects of our developed communication metrics show positive correlations between the measures we developed and performance on the synthetic task, and further research in this area could prove beneficial. Another promising aspect is that, even though we had to perform many of the counts and measures by hand, the same measures would not be very difficult to automate, decreasing the evaluation time significantly.

The single biggest hindrance to the efficiency of the communication evaluation is the incapability of the voice recognition software, which ultimately proved to be useless. However, the voice recognition software was off-the-shelf software, and better, more accurate software could possibly help with this problem. If the technology improves to the extent of being able to...
accurately transcribe communication to a word document, voice recognition software would be extremely useful.

Analyzing verbal protocol data in order to understand team performance can take a significant amount of time. Sanderson and Fisher (1994) report that some analysis techniques take as much as 10 hours for every hour of communication data. Our analysis of communication took approximately 2.5 hours per 15 minutes of communication data. Therefore, our analysis would benefit from a method where the processing time could be reduced and still obtain an accurate representation of the team process. When voice recognition software improves, it could turn into a viable option for effectively and efficiently evaluating team communication. For now, the method has to wait for the technology to catch up, but as soon as technology becomes sufficient, it should definitely be explored further.

REFERENCES


Cross-National Examination of the Aspects of Command

Stanley M. Halpin
Chief, Leader Development Research Unit
U. S. Army Research Institute
851 McClellan Avenue, Fort Leavenworth KS 66027
Stanley.Halpin@leavenworth.army.mil
Carol McCann
Ross Pigeau
Defence Research and Development Canada – Toronto
1133 Sheppard Avenue W., Toronto, Ontario M3M 3B9
Carol.McCann@drdc-rddc.gc.ca, Ross.Pigeau@drdc-rddc.gc.ca

ABSTRACT

In a series of papers, Carol McCann and Ross Pigeau have developed a new conceptualization of military command and control (C^2). They emphasize the importance of Command which they define as: “The establishment of common intent to achieve coordinated action.” McCann and Pigeau have identified three capabilities or aspects of Command: Competence, Authority, and Responsibility. An instrument, referred to as the Aspects of Command Survey, has been developed to assess the perceptions of military personnel of these dimensions. Early results from Canadian personnel showed striking regularities in the perceived relationship among these attributes of command as a function of rank. These data are consistent with a McCann & Pigeau hypothesis of a balance between the dimensions called the “balanced command envelope.”

In the present paper, Aspects of Command Survey data collected from U.S. Army personnel is presented and discussed. This preliminary analysis, based on data from 422 U.S. Army Company-grade officers, will examine differences among these Second Lieutenants, First Lieutenants, and Captains. We also compare the153 U. S. Army Captains with 51 Canadian officers. Of particular interest is the extent to which similar pattern-shifts occur across respondents as a function of rank. The discussion of similarities and differences across samples will further explore the implications for effective communications and shared understanding within multinational headquarters or other situations involving allied forces.

INTRODUCTION

This paper weaves together two threads from ongoing research on leadership and command. On the one hand we examine preliminary data that has been collected from U.S. Army and Canadian land force officers in an exploration of perceptions of the
dimensions of Command. This work derives from recent theoretical work by Carol McCann and Ross Pigeau on the nature of Command and Control of military forces. The second thread is the thesis that culture, and cultural differences, impact on the ability of cross-national teams to perform effectively. As our respective nations jointly and separately are often engaged in peacekeeping and other military missions as members of NATO and other alliances, it is of obvious concern to explore possible cultural differences (of Command) and to identify approaches for minimizing or eliminating the negative effects that such cultural differences may entail. This paper is the beginning of a project to explore the notion that core military cultural values, such as soldiers’ understanding of the meaning of “Command”, may be held in common across many nations’ military forces, and that the impact of such common core values and beliefs will supersede the impact of national cultural differences. In the current paper we first briefly review the tenets of the Pigeau-McCann framework for Command. We then examine a set of data gathered from U.S. Army officers using a survey instrument developed by McCann and Pigeau. We will discuss both the findings concerning the instrument itself and the findings as they relate to the underlying theoretical framework. We then compare the U.S. data with data obtained from a similar but smaller sample of Canadian officers, and discuss the possible implications of the identified similarities and differences.

PIGEAU-MCCANN FRAMEWORK FOR COMMAND

Following an extensive review of both military Command and Control (C²) doctrine and the scientific literature, Pigeau and McCann (1995) and McCann and Pigeau (1996) determined that there is a weak conceptual base for research in this area. Existing concepts of Command, Control and C² – as well as the relationship between these concepts – are poorly articulated and emphasise the role of technology as opposed to the role of the human. Indeed, they noted that “…the heavy emphasis on Control activities (e.g., procedures, algorithms, systems, etc) … undermined the critical role of Command.” (McCann and Pigeau, 1999). Pigeau and McCann have proposed a new human-centric framework for C² centered on the concept of Command, with Control in a supporting role. In contrast to the widely-referenced NATO definition of Command that focuses solely on authority, they have re-defined the concept of Command as “the creative expression of human will necessary to accomplish the mission” (Pigeau and McCann, 2002). Notwithstanding that Command has normally been associated with the position of Commander in military writing, the Pigeau-McCann framework does not limit the action of Command to those individuals in command positions -- as their definition makes clear. Thus they assert that any person can, in principle, command. This broadens the application of the concept to all people in the military, regardless of rank and position. In the framework, they further identify three capabilities that are essential for Command: competency, authority, and responsibility. The most extensive discussion of the relevant portions of the framework, summarized in the remainder of this section, is in Pigeau and McCann (2002).

With respect to competency, they distinguish four principle types: physical, intellectual, emotional, and interpersonal. A combination of the appropriate skill sets and abilities in these four areas would enable a person to perform the tasks associated with Command (e.g., interpreting the tactical situation, coping with the stress of an operation,
interacting with subordinates), but are not, of themselves, sufficient for effective Command.

Command also entails authority, and they draw a distinction between legal authority and personal authority. Legal authority derives from some external source of legitimacy, typically a military person’s government. Within a military, historical traditions and local unit-level customs may modify the degree to which, for example, legal authority for disciplinary action is delegated from higher to lower Command positions. Despite these variations, every regular military force may be expected to endow all Command positions with the authority to accomplish the required missions. Personal authority, on the other hand, flows from peers and subordinates. This authority is acknowledged at a tacit level, and is typically established by an individual over time “…based on reputation, experience, and character, and is often acquired through personal example . . .” (Pigeau and McCann, 2002). They further speculate that an individual’s personal authority is correlated with professional knowledge (i.e., competency) but is primarily based on an individual’s ethics, values, courage, and integrity (presumably, as these characteristics are perceived by others).

The third aspect of Command in the Pigeau-McCann model is responsibility. Responsibility may be viewed as either extrinsic or intrinsic. Extrinsic responsibility relates to the formal obligation to answer publicly for actions taken with legal authority. Extrinsic responsibility is thus the willingness to be accountable. Intrinsic responsibility, on the other hand, is self-generated. As described by Pigeau and McCann (2002) it is “…associated with the concepts of honor, loyalty, and duty.” They speculate that intrinsic responsibility is a function of the “…resolve and motivation an individual brings to a problem – the amount of ownership taken and the amount of commitment expressed.”

The framework hypothesizes that any person’s Command capability can be placed within a three-dimensional Command capability space (Pigeau and McCann, 2002). Furthermore, it asserts that for Command to be effective, the three aspects of Command (i.e., competency, authority and responsibility) should be commensurate with one another: that the aspects should be in a linear relationship called the Balanced Command Envelope (BCE). Thus, as a new recruit to the military gains competency, his or her authority to act (initially low) should increase in step, as should his or her responsibility. At the other end of the spectrum, a senior leader who has been given extensive authority should have a high degree of competency to wield that authority effectively – and should demonstrate a commensurately high level of responsibility. A mismatch between the aspects will compromise command effectiveness in different ways as described in McCann and Pigeau (1999).

One of the attractive characteristics of the CAR model of Command capability is that it provides the richness necessary to characterize this very complex human activity. The three aspects of Command, i.e., competency, authority, and responsibility, provide a deceptively simple structure. The first element, competency, is clearly and specifically focused on the individual’s capability -- the various key skills and knowledge that an individual brings to a Command setting. Authority, however, has sub-elements which derive both from an external system (legal authority) and from external perceptions of the individual (personal authority) which are in turn thought to be based in part on characteristics of the individual. The third element, responsibility, has sub-elements related to an external system (extrinsic responsibility, or accountability) and to
characteristics of the individual such as honor and loyalty. Integrating the three dimensions is the concept of a Balanced Command Envelope.

ASPECTS OF COMMAND INSTRUMENT

The instrument which was developed to explore the relationship between competency, authority and responsibility is referred to as the Aspects of Command Survey (AOCS). The eight sub-elements of the aspects of command, e.g., physical competency, extrinsic responsibility, etc., are listed separately in the AOCS. The respondent is asked to consider each of these sub-elements and, for each rank within his or her military, is asked to specify the current actual and ideal levels of the aspect. Thus, starting with Privates and proceeding to General Officers, the respondent indicates on a seven-point Low-to-High scale his or her perception of the actual and ideal levels on each of the sub-elements. Since there is not a perfect correspondence in the rank structure across nations (or even across services within nations), some modification to the instrument is needed for the intended military participants. There is a close correspondence in the naming of officer ranks between the U.S. army and the Canadian land force. However, this is not the case between the U.S. enlisted ranks and the Canadian Non-Commissioned Member (NCM) ranks. Notwithstanding the difference in rank names, there is a rough comparability in training and experience as one moves up the enlisted-NCM rank structure for both nations. Table 1 shows the correspondence of the target ranks between U.S. and Canadian armies that was made for this study. Note that U.S. WOs and CWOs have no direct equivalent in the Canadian Forces: e.g., the Canadian WO is equivalent to the U.S. Sergeant First Class. Note also that the AOCS instrument does not distinguish between 2LT’s and 1LT’s, but rather has one LT’s rank to be judged.

TABLE 1: U.S. - CANADIAN TARGET RANK CORRESPONDENCES

<table>
<thead>
<tr>
<th>US ARMY</th>
<th>CANADIAN ARMY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private (Pvt)</td>
<td>Private</td>
</tr>
<tr>
<td>Private First Class (PFC)</td>
<td>Corporal</td>
</tr>
<tr>
<td>Specialist (Spec)</td>
<td>Master Corporal</td>
</tr>
<tr>
<td>Sergeant (Sgt)</td>
<td>Sergeant</td>
</tr>
<tr>
<td>Staff Sergeant (SSG)</td>
<td>Sergeant Major (SGM)</td>
</tr>
<tr>
<td>Sergeant First Class (SFC)</td>
<td>Warrant Officer</td>
</tr>
<tr>
<td>Master Sergeant (MSG)</td>
<td>Master Warrant Officer</td>
</tr>
<tr>
<td>Staff Sergeant Major (SGM)</td>
<td>Chief Warrant Officer</td>
</tr>
<tr>
<td>Warrant Officer (WO)</td>
<td></td>
</tr>
<tr>
<td>Chief Warrant Officer (CWO)</td>
<td></td>
</tr>
<tr>
<td>Lieutenant (LT)</td>
<td>Lieutenant</td>
</tr>
<tr>
<td>Captain (CPT)</td>
<td>Captain</td>
</tr>
<tr>
<td>Major (MAJ)</td>
<td>Major</td>
</tr>
<tr>
<td>Lieutenant Colonel (LTC)</td>
<td>Lieutenant Colonel</td>
</tr>
<tr>
<td>Colonel (COL)</td>
<td>Colonel</td>
</tr>
<tr>
<td>Brigadier General (BG)</td>
<td>Brigadier General</td>
</tr>
</tbody>
</table>
DATA COLLECTION METHOD

The AOCS was administered in small group settings in both the U.S. and Canadian armies. Within the U.S. Army, we had access during this year to six different installations wherein we were able to arrange for data collection from groups of company grade officers (lieutenants and captains). Those eligible persons who agreed to participate came to a central location at the end of a work day, were provided with a very brief overview of the research activity, and proceeded with minimal spoken instruction to completion of the AOCS instrument. Fifteen to twenty-five persons might participate on a typical afternoon. Note that one or two other research projects were being conducted during the main part of the day, and some but not all of our sample of officers participated in multiple projects. In this sample, 429 active component officers completed the paper-and-pencil self-administrated Aspects of Command survey instrument. The majority of the respondents were male (90%), and the average respondent age was 27 (age range: 22-41). Most were at the rank of 1LT (189) and CPT (153). The most frequent duty position reported was PLT LDR (23%), and 30% of the respondents reported participation in zero operations.

Data collection in Canada was conducted in a similar manner at two locations: the Canadian Land Force Command and Staff College and the Royal Canadian Dragoons at Canadian Forces Base Petawawa for a total of 51 respondents. This data collection was completed in early 2001. All but two respondents were male, with an average age of 34 years (range: 29-42). Thirty-six were at the rank of Capt and the remainder were Majors. Most Canadian respondents (85%) reported at least one operational deployment.

DATA ANALYSES

For ease of presentation we aggregated the judgments of the sub-aspects for each command dimension – for example, by taking the mean value of the judgments of extrinsic and intrinsic responsibility. Analyses of the scale reliabilities were performed and showed that for the U.S. sample, alphas were .90 or higher for all sub-scales and for the aggregated competency, authority, and responsibility scales. The smaller Canadian sample yielded reliabilities somewhat lower, but still in the .80 range and higher. A comparison of the ratings by officers of different rank showed that the U.S. sub-samples (consisting of the ranks of 2LT, 1LT, and CPT) were significantly different in their ratings on all command dimensions, with the exception of Ideal responsibility. In this brief paper we will not discuss these differences in any detail. A visual inspection of the data suggests that these differences were primarily due to a tendency on the part of 2LT’s to rate LT’s somewhat higher (in terms of both Actual and Ideal levels of authority, competency and in terms of Actual responsibility) than they were rated by 1LT’s and/or CPT’s. Within the Canadian sample, there were no significant differences between CPT’s and MAJ’s, and those two data sets were combined. Since this paper is primarily

26 Within the U.S. Army, several installations within the continental United States are requested to set aside one week each year for the purpose of supporting data collection needs of a range of research agencies.
27 Note that physical competency was not included in the mean competency calculation because respondents rated both Ideal and Actual levels of physical competency as decreasing across the target ranks. The other competencies (e.g., intellectual, emotional and interpersonal) were rated as increasing with target ranks. The manner in which the different sub-aspects of the competency dimension should be combined to yield a single competency value is still under study. For simplicity sake, therefore, the mean competency calculation used in this paper includes only three of the four competencies.
Concerned with cross-national differences on command dimensions, all subsequent analyses will compare the U.S. CPT data to the combined Canadian data (i.e., CPT’s and MAJ’s).

Recall that the Pigeau-McCann perspective on Command would suggest that effective Command would be best achieved within a “Balanced Command Envelope” (BCE) – that is, the three aspects of authority, responsibility, and competency should be roughly aligned, or balanced. Therefore, as one examines higher ranks within the military command structure, one would expect higher levels of authority, responsibility, and competency. Two methods were used to demonstrate this relationship: 3-D scatterplots of the data, and mean plots for each of the command dimensions across target ranks.

As can be seen in the 3-D plots below (Figures 1a & 1b), both the U.S. and the Canadian judgments of the Actual aspects demonstrate a roughly linear BCE extending from lower levels of competency, authority and responsibility (values of 2 or 3) to higher levels (values of 5 or 6). While outliers do exist, the data tend to cluster along the diagonal of the 3-dimensional command capability space. Figures 1a & 1b illustrate the results of the Actual judgments. The 3-D scatterplots for the Ideal judgments were similar except that they were generally higher (i.e., they were also centered on the diagonal but were shifted towards the upper end of the range). Figures 1a & 1b illustrate that, in general, ratings of competency, authority and responsibility are correlated within the samples, as would be predicted by the BCE.

Three dimensional scatterplots are useful for illustrating general trends, but to explore specific differences between nations on Actual and Ideal judgments, more sensitive analyses are needed. Figures 2a, 2c and 2e show the mean values for competency, authority and responsibility as plotted against targets ranks for both the U.S. and Canadian samples. Notice that, for the most part, the shape of the curves for the Ideal judgments are similar for the U.S. and the Canadian samples. Judgments for Ideal levels of competency, authority and responsibility rise from the Private to the Sergeant Major, fall off for the junior officer (roughly to the level of the SSG) and then rise again to the Brigadier General. These results are consistent with the BCE hypothesis.
Although ANOVAs reveal statistically significant main effects between U.S. and Canadian Ideal judgments for all three command dimensions, the differences are minor compared to those observed between the U.S. and Canada on Actual judgments (see the closed symbol plots in Figures 2a, 2c and 2e). For many of the target ranks, Actual judgments for U.S. CPTs for each of the command dimensions match or exceed the judgments of Canadian CPTs and MAJs. The one exception is where Canadians consistently rate their privates higher than the U.S. respondents rate their privates.

It should be remembered, however, that since the U.S. and Canadian respondents were asked to make judgments for their nation only – i.e., neither nation was asked to assess the other’s target ranks – direct comparisons between the two nations on this measure are questionable. There is no assurance, for example, that the U.S. and Canadian respondents used the scale in a similar manner when making their judgments. To allow a more direct comparison, difference scores between Actual and Ideal judgments were calculated, on the assumption that each respondent at least used the scale consistently when making his or her own judgments. This strategy essentially anchors each respondent’s Actual judgment against his or her Ideal, regardless of their nationality, yielding a difference score that is more amenable to cross-nation comparison.

Although the Ideal judgments for the Canadian and U.S. respondents look remarkably similar (see Figures 2a, 2c and 2e), recall that statistically significant differences were observed.
Figures 2b, 2d and 2f illustrate the mean difference scores for each nation on competency, authority and responsibility across the target ranks. Zero represents Ideal judgments while negative deviations from zero represent the degree to which Actual judgments differed from Ideal judgments. For estimates of competency (Figure 2b), Canadian CPTs and MAJs judge their Privates, Corporals and Master Corporals to be closer to ideals than U.S. CPTs judge their Privates, Private First Class, Specialists and Sergeants. A cross-over occurs at the rank of Staff Sergeant and Sergeant First Class (or the Canadian equivalent: Sergeant and Warrant Officer) with the U.S. surpassing the Canadians at the Master Sergeant and Staff Sergeant Major levels (or the Canadian Master and Chief Warrant Officer levels). The U.S. and Canadian results are roughly similar for junior officer ranks but diverge sharply at more senior levels. U.S. Lieutenant Colonels, Colonels and Brigadiers General are perceived to be progressively closer to Ideal estimates whereas Canadian senior officers are perceived to be the same distance from Ideal levels. Difference scores for authority (Figure 2d) show that U.S. and Canadian enlisted/NCM’s are perceived roughly equivalently (except perhaps Sergeants and Sergeants First Class who seemed to diverge more from Ideal estimates). Senior officer ranks, however, show a pattern similar to that found for competency: U.S. senior officers are perceived as having levels of authority closer to Ideal levels whereas Canadian senior officers are perceived as having levels of authority quite removed from Ideal levels. Finally, difference scores for responsibility (Figure 2f) show two interesting findings. First, overall, estimates of responsibility are the furthest away from Ideal levels (compared to those found for competency and authority). This is particularly the case for junior non-commissioned members and Lieutenants. Second, the divergence seen between senior officers on the competency and authority dimensions is even more pronounced here. Canadian Colonels and Brigadier Generals are perceived to lie even further away from ideals than Canadian Captains, Majors and Lieutenant Colonels.

DISCUSSION

The preliminary nature of this study cautions us not to draw firm conclusions concerning the Command capability of those in the U.S. and Canadian armies. The findings, though intriguing, could be due to differences between the two nations’ militaries that are unrelated to Command. For instance, the U.S. army is more...
than an order of magnitude larger than the Canadian army. To what extent are its members exposed to the full range of target ranks? Is the small size of the Canadian army actually an advantage because its members stand a better chance of being exposed to the full range of army ranks? Or is it a disadvantage because a single poor example of competency, authority and responsibility stands an increased chance of reverberating throughout the army’s population? Regardless, the Pigeau-McCann approach to command taken in this paper offers a systematic way of quantifying Command capability. It offers a coherent taxonomy for labeling aspects of Command, one that can at least be used to alert potential areas of concern to military leaders – e.g., the large discrepancy between Ideal and Actual judgments of competency among U.S. junior enlisted members, or the very large discrepancy between Ideal and Actual judgments of responsibility among senior Canadian officers. Given the analyses which have been conducted, we have provided support for both the instrument and the theory of Command on which it was based. Further refinement of both is warranted and worthwhile.

A second interest in this research, beyond Command itself, is in the commonality of perception of Command across militaries. While the U.S. and Canadian forces are in many respects very much alike, it is also true that strong cultural differences exist between our two nations. Thus it is a promising outcome of this study that Ideal estimates of competency, authority and responsibility are strongly shared between the two military cultures. Any differences that do exist between nations on Ideal judgments are substantially smaller that differences between nations on Actual judgments, suggesting greater agreement for what the Ideal levels should be. In further work we will look at means for identifying additional shared beliefs and develop approaches to effective team work which take advantage of such commonalities.

REFERENCES


AN APPRAISAL OF A DEFENSE FINANCE SYMPOSIUM

Raymond O. Waldkoetter, Ed.D.
Educational and Selection Programs
Greenwood, IN 46143 USA
staff2@officeassistanceandsupplies.com

INTRODUCTION

The purpose of this study was to make a provisional appraisal of a Defense Finance Customer Service Symposium presented during June 2000. Customer service is regarded as an overall function of Defense Finance where the transacting needs of employees and users both internally and externally, and those of vendors, are addressed and satisfied according to the types of services provided through the combined organizational structure of the related finance operations. This structure includes categories of service from the specific actions such as civilian and military pay, given accounting operations, and payment, to paying vendors for products and support received. The symposium was convened to give in-service training for participants (customers) to maintain and improve the services that the participants rendered and supervised for those depending on and receiving the expected transactions.

A rationale for the Defense Finance Symposium was seen as an opportunity to enhance self-efficacy, which is belief in one’s abilities and skills to organize and apply the courses-of-action required to achieve given work objectives (Bandura, 1997). Likewise, collective efficacy also would be enhanced as a shared belief at group level in common abilities and skills to organize and apply the courses-of-action required to achieve given levels for the chosen work objectives. Respectively, the higher one’s self-efficacy is the more likely an individual is to perform and persist in task related behavior, whereas collective efficacy can offer positive group motivation and performance. Efficacy beliefs then at the individual and group levels can bring about important individual and organizational effects or results (Chen & Bliese, 2002).

Since efficacy beliefs can account for critical organizational results, a better understanding of how organizations can enhance self-and collective efficacy is important. Here a corollary purpose of this appraisal study was to examine whether certain organizational indicators of self-and collective efficacy are sufficiently specific and useful to be applied to positively affect the appraisal of self-and collective efficacy in the Defense Customer Service environment. In this appraisal key indicators of potential efficacy have been selected to make some determination of how well the in-service training has been received by the participants and to what extent responses can be used to further maintain and improve the delivery and performance of the related financial service operations. It was also assumed in this organizational context that participants working in the same work groups are likely affected by similar management behaviors (Shamir, Zakay, Breinin, & Popper, 1998).

METHOD
The symposium appraisal form was administered shortly before the end of the program to the 300 attending participants representing a substantial cross-section of associated financial operations and organizations. Participants were widely experienced and qualified to give an assessment of the general program areas and workshop topics. Participants were asked to complete the appraisal rating form of numerically scaled items for six general areas and nine separate workshops using a five-point rating scale, very poor (1), poor (2), okay (3), good (4), and very good (5). In other studies involving program evaluation, self-and collective efficacy rating forms have been reliably utilized (Jex & Bliese, 1999; Marlowe, 1986).

Six general areas specified for the symposium program were guest speakers, effect of general sessions, relevance, food and lodging, program organization, and overall rating. Nine separate workshops were defined as accounts payable initiatives, property reporting, Defense joint accounting system, Defense corporate information infrastructure, single stock fund, obligated practices/problem disbursing, year-end closing, paperless initiatives, and central disbursement update. Open-ended comments were solicited from participants regarding the symposium’s major strengths and weaknesses, and suggestions for improvement, with nominal groupings identified by a panel of three for the elicited comments within the listings collected, according to major strengths and weaknesses, and suggestions for improvement.

It is posited that within the administration and use of the symposium’s appraisal form and assessments obtained, the principal elements of self-and collective efficacy are reasonably incorporated as efficacy beliefs in terms of managerial climate, work experience, role clarity, and psychological strain or stress (Derogatis & Melisaratos, 1983). Accordingly positive ratings and comments, negative comments, and suggestions for improvement will be noted and discussed in regard to those general areas’ and workshops’ ratings, major strengths and weaknesses, and suggestions for improvement submitted by symposium participants. Previous research has indicated the best correlate of self-efficacy was the individual’s sense of role clarity with psychological strain/stress and work experience following, respectively. However, the best predictor of collective efficacy appeared to be a unit’s or section’s joint perception of upper-level management (Chen & Bliese, 2002). The above facets of this study’s appraisal are integrated and presented with added credence inferred as the following analysis and explanation are offered.

RESULTS AND DISCUSSION

Symposium appraisal forms were returned and each extensively reviewed to analyze participant ratings and responses with a random sample (N=46) of 15.3% drawn, reaching the 95% confidence level, used as a representative overview. It was a panel consensus a thorough interpretation of the symposium’s appraisal results would be derived with statistical validity from the above sample. Statistical testing using the one sample t-test showed ratings did not significantly differ (p>.05) from the averages (means) for the six general areas and nine separate workshops (SPSS, 1999). This would
suggest that any differences among ratings can be regarded as chance or random rather than describing any systematic intent by respondents. Some practical validity differences may be inferred as seen by the scope of positively and negatively elicited comments.

General areas’ ratings were quite positive and ranged from 3.98 (approximating good) for guest speakers to 4.35 (above good) for the symposium’s organization as seen in Table 1. Workshops’ ratings were also positive and ranged from 3.75 (okay to good) for accounts payable initiatives to 4.30 (above good) for central disbursing update. As with general areas the workshops received ratings close to or slightly over 4.0 (good).

Table 1
Symposium General Areas’ and Workshops’ Rating Averages (Means) *

<table>
<thead>
<tr>
<th>General Areas/Ratings</th>
<th>Workshops/Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guest speakers</td>
<td>Accounts payable initiatives</td>
</tr>
<tr>
<td>3.98</td>
<td>Property reporting</td>
</tr>
<tr>
<td>3.75</td>
<td></td>
</tr>
<tr>
<td>Effect of general sessions</td>
<td>Defense joint accounting system</td>
</tr>
<tr>
<td>4.00</td>
<td>Defense corporate information</td>
</tr>
<tr>
<td>3.95</td>
<td>infrastructure</td>
</tr>
<tr>
<td>Relevance</td>
<td>Single stock fund</td>
</tr>
<tr>
<td>4.00</td>
<td></td>
</tr>
<tr>
<td>Overall rating</td>
<td>Obligated practices/problem disbursing</td>
</tr>
<tr>
<td>4.00</td>
<td></td>
</tr>
<tr>
<td>Food/lodging</td>
<td>Year-end closing</td>
</tr>
<tr>
<td>4.30</td>
<td></td>
</tr>
<tr>
<td>4.06</td>
<td>Paperless initiatives</td>
</tr>
<tr>
<td>4.27</td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>Central disbursing update</td>
</tr>
<tr>
<td>4.30</td>
<td></td>
</tr>
<tr>
<td>Total Mean</td>
<td>Total Mean</td>
</tr>
<tr>
<td>4.04</td>
<td></td>
</tr>
</tbody>
</table>

* p > .05.

From the request for open-ended comments yielding the major strengths and weaknesses, and suggestions for improvement, and the nominal groupings as identified for the elicited comments, significantly 75% or more of the comments were briefly accounted for in each following comment grouping. For strengths work relevance,
workshop presenters, networking opportunities, guest speakers, and demonstrations with the associated professional organization accounted for this major observation. Weaknesses are principally noted as major negative miscellaneous workshop problems and comments, workshop scheduling, demonstrations, equipment-furniture and environment, and the agenda and a negligible military presence. Suggestions for improvement and related recommendations were listed as general scheduling, miscellaneous concerns and workshop recommendations, positive relevance to be emphasized, and participant/customer input needed and materials used.

As a symposium summary appraisal effects of the elements of self-efficacy and collective efficacy can be attributed to the general areas’ and workshops’ ratings, major strengths and weaknesses, and suggestions for improvement. In terms of self-efficacy role clarity, psychological strain/stress, work experience, and managerial climate were explored in view of the elicited ratings, strengths and weakness, and suggestions for improvement. These data have been reviewed subsequently for suitable application by the concerned echelons in the Defense Finance organization (DFAS, 2000). Collective efficacy though relating to positive role clarity, psychological strain/stress, and work experience appears to be more strongly related to the quality of managerial climate. Comments about managerial climate can be readily inferred as reviewers and management analysts evaluate the participant/customer responses to the form prior to any further appraisal (Katz & Kahn, 1978).

To the degree the data are utilized self-efficacy and collective efficacy may be enhanced by the democratic use of this appraisal for a Defense Finance Symposium. Even though various disparate or contrasting responses were received across the groupings of major strengths and weaknesses, and suggestions for improvement, they can be resolved when interpreted in the context of specific modifying statements provided by the participants in their expanded comments. While ratings for general areas and workshops were quite positive with a clearly positive overall symposium rating, invited open-ended comments proved sufficiently graphic to assist in the critical review and planning of future finance symposium programs.

REFERENCES


INTRODUCTION


CUSTOMER SERVICE OVERALL FUNCTION OF DEFENSE FINANCE WHERE TRANSACTING NEEDS OF EMPLOYEES/USERS (INTERNALLY AND EXTERNALLY) & OF VENDORS, ADDRESSED & SATISFIED ACCORDING TO TYPES OF SERVICES PROVIDED BY THE COMBINED STRUCTURE OF RELATED FINANCE OPERATIONS.

STRUCTURE INCLUDES CATEGORIES OF SERVICE FROM THE SPECIFIC ACTIONS SUCH AS CIVILIAN & MILITARY PAY, ACCOUNTING OPERATIONS & PAYMENT, TO PAYING VENDORS FOR PRODUCTS AND SUPPORT RECEIVED.

SYMPOSIUM CONVENED TO GIVE IN-SERVICE TRAINING FOR PARTICIPANTS (CUSTOMERS) TO MAINTAIN & IMPROVE SERVICES PARTICIPANTS RENDERED & SUPERVISED FOR THOSE DEPENDING ON & RECEIVING THE EXPECTED TRANSACTIONS.

RATIONALE FOR DEFENSE FINANCE SYMPOSIUM SEEN AS OPPORTUNITY TO ENHANCE SELF-EFFICACY-BELIEF IN ONE’S ABILITIES & SKILLS TO ORGANIZE & APPLY COURSES-OF-ACTION TO ACHIEVE GIVEN WORK OBJECTIVES (BANDURA,1997).

COLLECTIVE EFFICACY ALSO ENHANCED AS SHARED BELIEF AT GROUP LEVEL IN COMMON ABILITIES & SKILLS TO ORGANIZE & APPLY COURSES-OF-ACTION TO ACHIEVE GIVEN LEVELS FOR CHOSEN WORK OBJECTIVES.

RESPECTIVELY, HIGHER ONE’S SELF-EFFICACY IS THE MORE LIKELY INDIVIDUAL IS TO PERFORM & PRESIST IN TASK RELATED BEHAVIORS.
COLLECTIVE EFFICACY CAN OFFER POSITIVE GROUP MOTIVATION & PERFORMANCE.

Efficacy beliefs at individual & group levels can bring about important individual & organizational effects/results (Chen & Bleise, 2002).

Since efficacy beliefs can account for critical organizational results, better understanding of how organizations can enhance self- & collective efficacy is important.

Corollary purpose of this appraisal study was to examine if certain organizational indicators of self- & collective efficacy are sufficiently specific & useful to positively affect appraisal of self- & collective efficacy in the defense customer service environment.

In this appraisal, key indicators of potential efficacy selected to make some determination of how well in-service training was received by participants and to what extent responses can be used to further maintain & improve delivery & performance of the related financial service operations.

Assumed in this organizational context that participants working in same work groups are likely affected by similar management behaviors (Shamir, Zakay, Breinen, & Popper, 1998).

METHOD

Symposium appraisal form administered shortly before end of June 2000 program to 300 attending participants representing substantial cross-section of associated financial operations & organizations.

Participants were widely experienced & qualified to give assessment of the general program areas & workshop topics.

Participants asked to complete appraisal rating form of numerically scaled items for six general areas & nine separate workshops using a five-point rating scale, very poor (1), poor (2), okay (3), good (4), & very good (5).
SELF- & COLLECTIVE EFFICACY RATING FORMS RELIABLY USED IN OTHER STUDIES INVOLVING PROGRAM EVALUATION (JEX & BLIESE, 1999; MARLOWE, 1986).

SIX GENERAL AREAS SPECIFIED FOR SYMPOSIUM PROGRAM WERE GUEST SPEAKERS, EFFECT OF GENERAL SESSIONS, RELEVANCE, FOOD & LODGING, PROGRAM ORGANIZATION, & OVERALL RATING. NINE SEPARATE WORKSHOPS WERE DEFINED-ACCOUNTS PAYABLE INITIATIVES, PROPERTY REPORTING, DEFENSE JOINT ACCOUNTING SYSTEM, DEFENSE CORPORATE INFORMATION INFRASTRUCTURE, SINGLE STOCKFUND, OBLIGATED PRACTICES/PROBLEM DISBURSING, YEAR-END CLOSING, PAPERLESS INITIATIVES, & CENTRAL DISBURSING UPDATE.

OPEN-ENDED COMMENTS WERE SOLICITED FROM PARTICIPANTS REGARDING SYMPOSIUM’S MAJOR STRENGTHS & WEAKNESSES, & SUGGESTIONS FOR IMPROVEMENT.

NOMINAL GROUPINGS IDENTIFIED BY PANEL FOR ELICITED COMMENTS WITHIN LISTINGS COLLECTED ACCORDING TO MAJOR STRENGTHS & WEAKNESSES, & SUGGESTIONS FOR IMPROVEMENT.

POSITED THAT WITHIN ADMINISTRATION & USE OF SYMPOSIUM’S APPRAISAL FORM & ASSESSMENTS OBTAINED, PRINCIPAL ELEMENTS OF SELF & COLLECTIVE EFFICACY ARE REASONABLE INCORPORATED AS EFFICACY BELIEFS IN TERMS OF MANAGERIAL CLIMATE, WORK EXPERIENCE, ROLE CLARITY, & PSYCHOLOGICAL STRAIN OR STRESS (DEROGATIS & MELISARATOS, 1983).

POSITIVE RATINGS & COMMENTS, NEGATIVE COMMENTS, & SUGGESTIONS FOR IMPROVEMENT NOTED & DISCUSSED IN REGARD TO THOSE GENERAL AREAS’ & WORKSHOPS’ RATINGS, MAJOR STRENGTHS & WEAKNESSES, & SUGGESTIONS FOR IMPROVEMENT SUBMITTED BY SYMPOSIUM PARTICIPANTS.

PREVIOUS RESEARCH INDICATED BEST CORRELATE OF SELF-EFFICACY WAS INDIVIDUAL’S SENSE OF ROLE CLARITY WITH PSYCHOLOGICAL STRAIN/STRESS & WORK EXPERIENCE FOLLOWING RESPECTIVELY. BEST PREDICTOR OF COLLECTIVE EFFICACY APPEARED TO BE A UNIT’S OR SECTION’S JOINT PERCEPTION OF UPPER-LEVEL MANAGEMENT (CHEN & BLIESE, 2002).

ABOVE FACETS OF THIS STUDY’S APPRAISAL ARE INTEGRATED & PRESENTED IN FOLLOWING ANALYSIS & EXPLANATION.
RESULTS AND DISCUSSION

SYMPOSIUM APPRAISAL FORMS WERE RETURNED AND EACH REVIEWED TO ANALYZE PARTICIPANT RATINGS AND RESPONSES WITH RANDOM SAMPLE (N=46) 15.3% DRAWN GIVING A 95% CONFIDENCE LEVEL, USED AS A REPRESENTATIVE OVERVIEW.

CONSENSUS OF PANEL WAS SAMPLE WOULD OFFER VALID INTERPRETATION OF SYMPOSIUM’S APPRAISAL RESULTS.

STATISTICAL TESTING (ONE-SAMPLE $t$ – TEST) SHOWED RATINGS DID NOT SIGNIFICANTLY DIFFER ($p > .05$) FROM AVERAGES (MEANS) FOR THE SIX GENERAL AREAS & NINE SEPARATE WORKSHOPS (SPSS, 1999).

DIFFERENCES AMONG RATINGS CAN BE DUE TO CHANCE RATHER THAN ANY SYSTEMATIC INTENT BY RESPONDENTS. SOME PRACTICAL VALIDITY DIFFERENCES MAY BE INFERRED AS SEEN BY SCOPE OF POSITIVE & NEGATIVE COMMENTS ELICITED.

GENERAL AREAS’ RATINGS QUITE POSITIVE AND RANGED FROM 3.98 (APPROX. GOOD) FOR GUEST SPEAKERS TO 4.35 (ABOVE GOOD) FOR THE SYMPOSIUM’S ORGANIZATION (TABLE 1). WORKSHOPS’ RATINGS ALSO POSITIVE AND RANGED FROM 3.75 (OKAY TO GOOD) FOR ACCOUNTS PAYABLE INITIATIVES TO 4.30 (ABOVE GOOD) FOR CENTRAL DISBURSING UPDATE.

TABLE 1: SYMPOSIUM GENERAL AREAS’ & WORKSHOPS’ RATING AVERAGES (MEANS)*

<table>
<thead>
<tr>
<th>GENERAL AREAS/RATINGS</th>
<th>WORKSHOPS/RATINGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>GUEST SPEAKERS</td>
<td>3.98</td>
</tr>
<tr>
<td>EFFECT OF GENERAL SESSIONS</td>
<td>4.00</td>
</tr>
<tr>
<td>RELEVANCE</td>
<td>4.20</td>
</tr>
<tr>
<td>OVERALL RATING</td>
<td>4.22</td>
</tr>
<tr>
<td>INFRASTRUCTURE</td>
<td>4.00</td>
</tr>
<tr>
<td>FOOD &amp; LODGING</td>
<td>4.30</td>
</tr>
<tr>
<td>ACCOUNTS PAYABLE INITIATIVES</td>
<td>3.75</td>
</tr>
<tr>
<td>PROPERTY REPORTING</td>
<td>3.90</td>
</tr>
<tr>
<td>DEFENSE JOINT ACCOUNTING SYSTEM</td>
<td>3.95</td>
</tr>
<tr>
<td>DEFENSE CORPORATE INFORMATION</td>
<td></td>
</tr>
<tr>
<td>SINGLE STOCK FUND</td>
<td>4.00</td>
</tr>
<tr>
<td>OBLIGATED PRACTICES/PROBLEM DISBURSING</td>
<td>4.06</td>
</tr>
<tr>
<td>YEAR-END CLOSING</td>
<td>4.14</td>
</tr>
<tr>
<td>PAPERLESS INITIATIVES</td>
<td>4.27</td>
</tr>
<tr>
<td>CENTRAL DISBURSING UPDATE</td>
<td>4.30</td>
</tr>
<tr>
<td>ORGANIZATION</td>
<td>4.35</td>
</tr>
<tr>
<td>TOTAL MEAN</td>
<td>4.17</td>
</tr>
</tbody>
</table>

44th Annual Conference of the International Military Testing Association
Ottawa, Canada, 22-24 October 2002
OPEN-ENDED COMMENTS YIELDED MAJOR STRENGTHS & WEAKNESSES, & SUGGESTIONS FOR IMPROVEMENT AND FROM NOMINAL GROUPINGS IDENTIFIED FOR ELICITED COMMENTS, 75% OR MORE OF COMMENTS SIGNIFICANTLY ACCOUNTED FOR BRIEFLY IN EACH FOLLOWING COMMENT GROUPING.

FOR STRENGTHS: WORK RELEVANCE, WORKSHOP PRESENTERS, NETWORKING OPPORTUNITIES, GUEST SPEAKERS, AND DEMONSTRATIONS WITH THE ASSOCIATED PROFESSIONAL ORGANIZATION ACCOUNTED FOR THIS MAJOR OBSERVATION.

WEAKNESSES WERE PRINCIPALLY NOTED AS MAJOR NEGATIVE MISCELLANEOUS WORKSHOP PROBLEMS AND COMMENTS, WORKSHOP SCHEDULING, DEMONSTRATIONS, EQUIPMENT-FURNITURE & ENVIRONMENT, AND THE AGENDA WITH NEGLIBIBLE MILITARY PRESENCE.

SUGGESTIONS FOR IMPROVEMENT & RELATED RECOMMENDATIONS WERE LISTED AS GENERAL SCHEDULING, MISCELLANEOUS CONCERNS & WORKSHOP RECOMMENDATIONS, POSITIVE RELEVANCE TO BE EMPHASIZED, AND PARTICIPANT/CUSTOMER INPUT NEEDED & MATERIALS USED.

SYMPOSIUM SUMMARY APPRAISAL: EFFECTS OF THE ELEMENTS OF SELF-EFFICACY & COLLECTIVE EFFICACY CAN BE ATTRIBUTED TO GENERAL AREAS' & WORKSHOPS' RATINGS. MAJOR STRENGTHS & WEAKNESSES, AND SUGGESTIONS FOR IMPROVEMENT.


COLLECTIVE EFFICACY THOUGH RELATING TO POSITIVE ROLE CLARITY, PSYCHOLOGICAL STRAIN/STRESS, AND WORK EXPERIENCE APPEARS TO BE MORE STRONGLY RELATED TO THE QUALITY OF MANAGERIAL CLIMATE.
COMMENTS ABOUT MANAGERIAL CLIMATE CAN BE INFERRED AS REVIEWERS AND MANAGEMENT ANALYSTS EVALUATE PARTICIPANT (CUSTOMER) RESPONSES TO THE FORM PRIOR TO ANY FURTHER APPRAISAL (KATZ & KAHN, 1978).

TO THE DEGREE THE DATA ARE UTILIZED SELF-EFFICACY & COLLECTIVE EFFICACY MAY BE ENHANCED BY THE DEMOCRATIC USE OF THIS APPRAISAL FOR A DEFENSE FINANCE SYMPOSIUM.

VARIOUS DISPARATE OR CONTRASTING RESPONSES WERE RECEIVED ACROSS THE GROUPINGS OF MAJOR STRENGTHS & WEAKNESSES AND SUGGESTIONS FOR IMPROVEMENT.

DIFFERENCES CAN BE RESOLVED WHEN INTERPRETED IN THE CONTEXT OF SPECIFIC MODIFYING STATEMENTS PROVIDED BY PARTICIPANTS/CUSTOMERS IN THEIR EXPANDED COMMENTS.

RATINGS FOR GENERAL AREAS & WORKSHOPS QUITE POSITIVE WITH CLEARLY POSITIVE OVERALL SYMPOSIUM RATING.

INVITED OPEN-ENDED COMMENTS PROVED SUFFICIENTLY GRAPHIC TO ASSIST IN CRITICAL REVIEW AND PLANNING OF FUTURE FINANCE SYMPOSIUM PROGRAMS.
Quality of Life in the Canadian Forces:  
An Introduction to Definition and Measurement

David Jenkins  
Directorate of Quality of Life, Department of National Defence

Ottawa, October 2002

Abstract

This abridged version of *Quality of Life in the Canadian Forces: Toward a National Index for Members and their Loved Ones* provides an introduction to the conceptualisation and measurement of quality of life (QoL) inside and outside the Canadian Forces (CF). QoL is an extremely broad concept that may be approached in any number of ways. These approaches focus either on the conditions necessary for a good life (social approaches) or the actual experience of a good life (individual approaches). Similarly indicators of QoL may be subjective or objective and each has advantages and drawbacks over the other. It is argued that subjective, individual assessment of QoL is most appropriate because conditions that facilitate a good QoL vary so much person to person. However the amelioration of QoL necessarily depends upon the co-ordination of subjective and objective data. It is also emphasised that the true definition of QoL comes in the way that it is measured or operationalized. The advantages and disadvantages of assessing QoL globally or examining various domains or areas of life are discussed.

Definition

Quality of life refers to “1) the presence of conditions deemed necessary for a good life [and/or] 2) the practice of good living” (Veehoven 1997, p.2). There is widespread disagreement over what “the good life” entails however, and, although there is some degree of consensus as to what aspects of a person’s life should be included within the QoL construct (e.g. health, work, leisure) it is by no means free from such debate. Definitions of QoL range from being unspecific as in: “the degree to which the experience of an individual’s life satisfies that individual’s wants and needs (both physical and psychological)” (Rice, 1984) to very unspecific as in: “the product of
interplay among social, health, economic and environmental conditions which affect human and social development” (Ontario Social Development Council, 2001).

There are a number of terms that are more or less synonymous with QoL and others that may appear to be but are not. Living conditions are a synonym of the first definition of QoL. Welfare is a possible synonym but, in addition to having a negative connotation the term is often used to refer to the conditions that insure a minimally decent life like the Basic Needs approach to measuring QoL. Perceived quality of life, subjective wellbeing (SWB), quality of wellbeing (QWB) and wellbeing may be used synonymously with the second definition of QoL (QoL = quality of experience). Happiness and life satisfaction are close to being synonymous with the second definition although in the current psychological research a distinction is made between cognitive evaluation (satisfaction) and levels of positive and negative affect sometimes called happiness. In some cases happiness is used to refer to pleasant affect while in others (e.g. Veehoven, 1997) along with life satisfaction it denotes an overall evaluation of life. In this discussion life satisfaction will refer to a person’s overall evaluation of their life while happiness will simply refer to the experience of QoL in general.

Numerous Approaches

Because so many events play some role in determining the quality of a person’s life, QoL may be approached from any number of perspectives and levels of analysis. In some fashion or other all disciplines aim to examine or improve certain aspects of the QoL of certain groups of people, though some do so more holistically or explicitly than others. While astrology may offer some semblance of advice to the tender minded astronomy may aim to further human understanding, or give some notice as to impending destruction of the planet. A politician may be interested in how the QoL of the nation compares with that of other nations and look at such things as per capita gross national product and mean level of education. A general practitioner will be more interested in keeping patients free from disease and will likely encourage a healthy diet and level of exercise. A psychologist will likely focus on such things as coping strategies, maladaptive beliefs, and optimism. Still these are only a few of the more obvious and stereotypical approaches to QoL. A painter, philosopher or preacher will likely have quite
different ideas about what is critical to a good life. At the individual level of analysis each person takes a different approach to life, regardless of profession.

In the same way, each organisation takes a different approach to QoL. Each organisation offers different goods and services to clients and employees and asks for different things in return (money, a favourable review, time, energy etc.). Each organisation has a different mission statement and set of objectives that change periodically to reflect a changing society. As such, each organisation seeks to ameliorate only certain aspects of a person’s life as reflected either in its definition of QoL or its human resource management strategies. In the CF, QoL is confounded with morale, job satisfaction, daily lifestyle, and conditions of service (Dowden, 2000). Morale, job satisfaction and conditions of service refer to only one area of a member’s life: his or her quality of working life (QoWL). But the relationship between QoWL and quality of non-work life is reciprocal. In other words each affects the QoL experienced in the other (Tait, Padgett & Baldwin, 1989).

Although any number of organisations and departments within them may use the generic term QoL it is important to realise that they only aim to improve certain aspects of QoL and do so in large part to better meet organisational objectives. As such QoL is easily confused with the aims of specific organisations, departments and policies.

Conceptual and Operational Definitions

Along with being defined as either the experience of a good life or the conditions deemed necessary for a good life (Veehoven, 1997), definitions for QoL come in two other forms as well: the conceptual and the operational. While the conceptual definition is the ideal form of the term it may not suggest what techniques of measurement are appropriate. Defining QoL conceptually is simple, QoL means happiness or the lack thereof. Defining QoL operationally on the other hand is done by choosing indicators of QoL like survey items or objective statistics. Since QoL is such a broad concept only a very small proportion of the variables that go into determining the quality of a person’s life are taken into account by any given operational definition. It is important to understand that, although most documents on QoL measurement provide a unique conceptual definition of QoL, it is the indicators that have been chosen in the study and the means by which they are evaluated that truly define the work being done.
Objective versus Subjective Approaches

Approaches that use objective statistics to measuring QoL in a country or organisation, like the GDP and HDI, employ the definition of QoL as quality of living conditions. Social approaches such as these evaluate the extent to which the proper material conditions are present for happiness. For example, if someone says that the QoL in Ethiopia is poor they are likely making a generalisation as to the availability of commodities and services like food, health care and education, not saying that no one in the country is happy. More recently though, with the advent of attitude surveys, it has become possible to maintain focus on the subjective experience of the individual. Individual approaches like the Four Factor Model of subjective wellbeing (see Diener, et al, 1999) attempt to evaluate not just the extent to which the proper precursors of happiness are present but the extent to which people are in fact experiencing happiness.

Objective versus Subjective Indicators

Because the experience of QoL is illusive, variables that affect it are called QoL indicators. Indicators may be classified quite easily as either objective or subjective according to how dependent they are on personal opinion. While an individual’s satisfaction with the size of his or her house is a subjective indicator, the actual dimensions of the house are an objective indicator. As social approaches to QoL look at the material conditions that contribute a good life they tend to employ objective QoL indicators. Individual approaches on the other hand tend to employ subjective indicators. According to Argyle (1996) the correlation between objective and subjective measures, health for example, are generally positive but weak. Although QoL is in the end a subjective experience and objective indicators are generally considered to be surrogate measures of subjective ones, there are benefits and faults with both objective and subjective indicators of QoL.

Objective indicators are largely independent of an individual’s opinion. They are often quantitative, that is to say numerical figures such as the number of ice rinks in a community. Figures for objective indicators are available from any number of agencies or businesses. Statistics on such things as crime, income and health are routinely recorded.
and usually available at a nominal cost. The CF itself has a great wealth of information about its members and the services provided them.

Subjective indicators require some personal evaluation or attitude. Normally with subjective indicators an individual is asked to self-report, that is to report on certain aspects of his or her own life. Alternatively a person may be asked to evaluate certain aspects of another person’s life as in the case of a physician examining a patient. Subjective responses may be numeric representations of how an individual feels, as in the case of asking members to rate their overall QoL on a scale from one to ten, or qualitative when individuals describe their situation in their own words.

Substantial costs are incurred when assessing subjective indicators, in the writing, administration and collection of surveys and the recording and analysis of data. Additional costs are incurred when asking subjects to express their thoughts and feelings verbally in an open-ended qualitative format, on paper or vocally in an interview or focus group. Much time must be spent collecting information from a statistically significant size of sample and then in condensing it into a useful report. A great deal of information is inevitably lost.

In addition to being costly and time consuming there are numerous obstacles to be negotiated in subjective assessment. These hazards include low reliability, response effects, positive bias and adaptation. It is especially important when administering instruments to employees to consider subjects’ motives in responding. For example, are CF members truly dissatisfied with their pay or is this a result of the belief that the squeaky wheel will get the oil?

The decision of which indicators to choose and which domains of life to assess is a difficult one. When objective indicators are employed exclusively or when members of the population are not consulted about what is important to them, lists of indicators are incomplete and often very lengthy. “The list of possible objective indicators is endless—wealth, health, employment, education… rainfall, alcohol consumption, literacy, etc.” (Argyle 1996 p.18). Because of this, and because the relationship between most objective indicators and subjective experience is by no means a reliable one, individual, subjective assessment of QoL is preferable.
Even with subjective indicators, if one should attempt to include every activity an individual regularly engages in the list quickly grows too large. But with a subjective approach, the task of selecting indicators is a much more manageable one because it is possible to ask a respondent about their overall satisfaction or happiness with a major domain, or area of their life. As such one might decide to assess respondents’ overall satisfaction with their work, relationships, leisure and health. Alternatively one might decide to assess in depth a single domain that is of particular interest, so as to get a better picture of what specific aspects of that domain require improvement.

**Global versus Domain Quality of Life**

While domain QoL refers to the QoL in one area of life, global QoL refers to quality of life as a whole or an individual’s overall assessment of their life. Global QoL may be assessed quite easily by asking questions like “how satisfied are you with your life overall” (e.g. Satisfaction with Life Scale-5 Diener et al., 1985). It is inadvisable to attempt to assess (subjective) global QoL through the summation of domain scores however. Even with ongoing advances in statistical techniques it will not be possible to construct a model that describes the relative contribution of each domain to a person’s global QoL. The inescapable fact is that the importance of each aspect of life (and the resulting contribution to wellbeing as a whole) differs from one person to another. Domain scores lend some explanation for the global response but are more useful for indicating what areas warrant greater attention and resources. Both domain scores and any importance ratings for those domains, are more meaningful if viewed separately from one another.

**Conclusions**

The preceding was an introduction to the conceptualisation and measurement of QoL inside and outside the CF. QoL is an extremely broad term that may be approached in any number of ways. These approaches focus either on the conditions necessary for a good life (social approaches) or the experience of a good life (individual approaches). Similarly indicators may be classified as subjective or objective and each has advantages and drawbacks over the other. It is further emphasised that the true definition of QoL comes in the way that it is measured. A person’s happiness with their life overall may be measured (global QoL) or areas of a person’s life may be investigated (domain QoL).
Assessing global QoL lends little insight into what areas of a person’s life need improvement but with a list of domains it is difficult to determine the relative importance of those domains. Individual, subjective assessment is preferable to objective assessment in part because the conditions that facilitate a good QoL vary so much between individuals and in part because describing QoL using objective figures produces an extremely lengthy and inevitably incomplete list of indicators. This is not to say that objective indicators of QoL are dispensable. The amelioration of QoL necessarily depends upon establishing links between, and co-ordinating objective and subjective data.

References


Dowden, C. (2000) Quality of Life in the Canadian Forces: Conceptualization and measurement of the QoL construct, (sponsor research report 00-05). Ottawa: DHRRE.

Dowden, C. (2001). Quality of Life in the Canadian Forces: Results from the National Survey. (sponsor research report 01-13) Ottawa: DHRRE.


Test Evaluation for Augmenting the Armed Services Vocational Aptitude Battery

Janet D. Held and Geoffrey E. Fedak
Navy Personnel Research, Studies, and Technology

Michael P. Crookenden and Thomas A. Blanco
Electronic Data Systems - Federal

The Department of Defense and the United States military branches individually are continually looking for new predictors of training and job performance to include in their military classification processes. The Armed Services Vocational Aptitude Battery (ASVAB) is the primary selection and classification instrument for this purpose. Two tests, Coding Speed (a perceptual speed test) and Assembling Objects (a spatial ability test), have demonstrated potential for improving classification for the United States Navy. The Navy has made three types of assessments for Coding Speed and Assembling Objects for classification purposes: 1) incremental validity when added to composites of ASVAB tests for use in the prediction of Navy school grades, 2) gender and minority group score differences that would present score barriers to schools, and 3) classification efficiency, in terms of qualifying and assigning a larger percentage of the recruit population for Navy jobs. Positive results were found on all three assessments for both tests. The results are reported in this paper. The evaluation is from the perspective of service use, and does not include the test evaluation criteria of test reliability or differential item functioning. This is addressed, for ASVAB, at the Defense Data Manpower Center (DMDC) - Personnel Testing Division, the ASVAB test developer.

Validities of Composites

Validities, corrected for range restriction, were computed for ASVAB composites, some of which contained the Assembling Objects (AO) and Coding Speed (CS) tests, for thirty-two of the approximately ninety Navy ratings that had a sufficient number of students having AO scores. The other ASVAB tests are: General Science (GS), Arithmetic Reasoning (AR), Word Knowledge (WK), Paragraph Comprehension (PC), Auto and Shop Information (AS), Mathematics Knowledge (MK), Mechanical Comprehension (MC), and Electronics Information (EI). Tests were adaptive except for CS. Scores were originally theta based then standardized with a mean of 50 and a standard deviation of 10 with military applicant population scores used for the normative linear conversion. Navy composite scores are simply a summation of test standard scores. The Verbal (VE) test is a weighted combination of the CAT-ASVAB WK and PC tests. The Auto and Shop Information (AS) test is a weighted combination the Auto Information (AI) and Shop Information (SI) tests.

Validity analyses are used to determine which ASVAB composites are most predictive of school success. Because school performance grades are not available for all members of the recruit population, statistical procedures are used to “correct” the school sample validity to obtain a population validity estimate (Gulliksen, 1950; Lawley 1943). However, the statistical correction procedures to obtain the population validity estimates are not robust when sample sizes are small, and when the assumptions for performing the corrections are violated, especially at very stringent selection ratios (Held & Foley, 1996). As a consequence, a bootstrap validation method designed to deal with small sample inadequacies (Held, 1997) was used in this study.

29 The opinions expressed in this paper are those of the author, are not official, and do not necessarily reflect the views of the Navy Department.
The bootstrap is generally used to study the error distribution of a specific statistic of interest (Efron & Tibshirani, 1993). However, the bootstrap was used in this study to distribute small sample sampling errors and errors that result from violating the assumptions for performing the corrections (Lawley, 1943) more or less equally among the terms that are used to compute the corrected validity. Although an accurate assessment of the validity magnitude and the validity magnitude difference between composites cannot be made with this procedure if there are extremely small samples (N = 87, was the smallest sample in this study), identification of composites that are most valid, can. In this study, the validity comparison was of the median corrected validity from 1,000 bootstrapped corrected validities generated for each composite within each school.

Table 1 provides the validity results and rating (job) information.

### Table 1
Rating Descriptions and Composite Median Validities from Bootstrap Procedure

<table>
<thead>
<tr>
<th>Rating</th>
<th>Description</th>
<th>Best ASVAB</th>
<th>Best ASVAB + AO</th>
<th>Validity Dif.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aviation Boatsman’s Mate</td>
<td>Services hydraulics and arresting gear maintenance</td>
<td>VE+AR+MK+MC</td>
<td>AR+GS+AS+AO</td>
<td>.013</td>
</tr>
<tr>
<td>(Equipment) (N = 244)</td>
<td></td>
<td>(.626)</td>
<td>(.639)</td>
<td></td>
</tr>
<tr>
<td>Builder</td>
<td>Performs wood and concrete construction</td>
<td>AR+MC+AS</td>
<td>AR+MC+AO</td>
<td>.015</td>
</tr>
<tr>
<td>(N = 339)</td>
<td></td>
<td>(.628)</td>
<td>(.643)</td>
<td></td>
</tr>
<tr>
<td>Construction Mechanic</td>
<td>Services gasoline and diesel engines</td>
<td>AR+MC+AS</td>
<td>AR+GS+AS+AO</td>
<td>.013</td>
</tr>
<tr>
<td>(N = 260)</td>
<td></td>
<td>(.573)</td>
<td>(.586)</td>
<td></td>
</tr>
<tr>
<td>Fire Control Technician</td>
<td>Maintains weapon telemetry systems - submarines</td>
<td>AR+MK+EI+GS</td>
<td>AR+MK+EI+AO</td>
<td>.017</td>
</tr>
<tr>
<td>(N = 87)</td>
<td></td>
<td>(.694)</td>
<td>(.711)</td>
<td></td>
</tr>
<tr>
<td>Parachute Rigger</td>
<td>Rigs parachutes, maintains survival equipment</td>
<td>AR+MK+EI+GS</td>
<td>AR+MK+EI+AO</td>
<td>.020</td>
</tr>
<tr>
<td>(N = 293)</td>
<td></td>
<td>(.656)</td>
<td>(.678)</td>
<td></td>
</tr>
<tr>
<td>Quartermaster</td>
<td>Steers ship; logs compass readings, tides, bearings, etc.</td>
<td>VE+AR+MK+MC</td>
<td>VE+AR+MK+AO</td>
<td>.012</td>
</tr>
<tr>
<td>(N = 250)</td>
<td></td>
<td>(.750)</td>
<td>(.762)</td>
<td></td>
</tr>
<tr>
<td>Signalman</td>
<td>Operates assorted visual communications devices</td>
<td>AR+MK+EI+GS</td>
<td>AR+MK+EI+AO</td>
<td>.035</td>
</tr>
<tr>
<td>(N = 149)</td>
<td></td>
<td>(.542)</td>
<td>(.577)</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Validities are corrected for range restriction; differences were computed as best ASVAB plus AO composite minus the best ASVAB composite.

The validity analyses showed that seven of the 32 schools involved in the study could utilize ASVAB selector composites that included the AO test based upon an increment in validity ranging from .01 to .03. A .05 validity increment in generally associated with an improved school success rate of about two percent (Taylor & Russell, 1939).
Score Difference Analysis

Adverse impact on a minor group occurs when the test used for selection or classification does not predict with the same sensitivity the criterion outcome that is measured for success (for this study, school performance as measure by a final school grade). The minor sample sizes were too small in the seven school samples that demonstrated higher validity for an AO composite to conduct regression line slope and intercept difference analyses between the major and minor groups. However, one of the 32 samples, the sample for the Aviation Electronics Technician rating, was large enough to make gender and race comparisons. Consistent with a joint service study led by DMDC conducted in 1992 to address this issue (Wise et al., 1992), the composites compared were determined to be equally sensitive for major and minor groups in the area of the relevant range of minimum qualifying scores. However, it is evident that score barriers do exist for certain groups on technical composites. For gender, for instance, females scored lower than males on the final school grade for the AT course (although by just one point) and significantly lower on the AR+MK+EI+GS and VE+AR+MK+MC composites. However, females scored higher than males on the VE+AR+MK+AO composite.

In order understand minor group score barriers for job classification and the contribution that high test scores on minor group neutral tests can make when used in the compensatory test composite model (scores are used additively without individual test minimum qualifying scores); an effect size analysis was conducted for a Navy recruit population in a year where AO scores first became available.

Table 2 lists effect sizes calculated as the major and minor group mean test score difference divided by the combined group test score standard deviation. Comparisons were made for races within gender to establish cultural or opportunity differences, and for genders (combining all races within each gender) to understand educational and general interest differences.

The same effect sizes patterns emerged for the race/ethnic groups for each gender, suggesting cultural differences. Effect sizes greater than .5 are discussed here. African Americans had the largest number of effect size differences, followed by Hispanics, and then Asians. No effect size differences were found for American Indians for either males or females. Not considering American Indians further, Auto Shop had the largest effect size (favoring Caucasians) for all major and minor group comparisons. Mathematics Knowledge had the only negative effect size, which was for Asians (favoring Asians), but not quite .5. Coding Speed had low effect sizes for all comparisons. Assembling Objects had greater then .5 effect size for only African Americans.

Effect sizes comparing the male major group and female minor group (males, N = 35,831; females, N = 8,246) combining all races are as follows in descending order with a positive effect size denoting higher scores for males: AS 1.05; MC .80, EI .74; GS .37; AR .36; VE .05; AO .15; MK -.15, CS -.38. Assembling Objects, a potential substitute for Auto Shop, had a marginal gender effect size, whereas Auto Shop had the highest gender effect size (favoring males). Mathematics Knowledge was one of two tests that favored females. Coding Speed, which is presently used in one Navy composite that functions as an alternative classification standard, had the highest effect size favoring women. Coding Speed, however, was not evaluated as optimal in the technically oriented ratings analyzed with AO in this study, and was used only in the classification efficiency simulation for ratings where incremental validity has been demonstrated.
Table 2
Effect Size Analysis for Subgroups From
The Recruit Accession Population

<table>
<thead>
<tr>
<th>Male Effect Sizes</th>
<th>Female Effect Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cauc. Group N = 22,230</td>
</tr>
<tr>
<td>ASVAB</td>
<td>Af.Am.</td>
</tr>
<tr>
<td></td>
<td>Hisp.</td>
</tr>
<tr>
<td></td>
<td>Asian</td>
</tr>
<tr>
<td></td>
<td>A. Ind.</td>
</tr>
<tr>
<td></td>
<td>Af.Am.</td>
</tr>
<tr>
<td></td>
<td>Hisp.</td>
</tr>
<tr>
<td></td>
<td>Asian</td>
</tr>
<tr>
<td></td>
<td>A. Ind.</td>
</tr>
</tbody>
</table>

*Denotes an effect size greater than .5 (standard deviation). A positive value indicates the major group had a higher mean.
Classification Efficiency Simulation

Three composite sets were formulated that included existing Navy operational composites and the AO and CS composites resulting from the validity analyses. Composite Set 1 (baseline) contained no composites with Coding Speed and Assembling Objects. Composite Set 2 contained several composites with AO. Composite Set 3 contained several composites with CS, and Composite set 4 contained the all of the composites with either CS or AO.

Classification efficiency for this study was defined as (1) the increase in the percentage of a recruit population classified to jobs and (2) the standard deviation of the fill rate among different types of jobs using augmented composite sets. Because the validity of the augmented composite sets are at least as high as the baseline composite set, the expected school performance outcomes (mean predicted performance) are at least as high as classification based upon the baseline composite set. Each composite set was used to establish the ASVAB composite and minimum qualifying score for the over 80 Navy ratings (and over 150 rating/advanced program combinations). Minimum qualifying scores were set for each rating under the four composite set conditions to result in the same recruit qualification rate.

The classification algorithm was developed by the Navy Personnel Research, Studies, and Technology (Folchi & Watson, 1997) and operationalized by EDS Federal Engineering and Logistics, Fort Worth, Texas under contract to the Navy Personnel Research, Studies, and Technology (NPRST) Division of the Navy Personnel Command, Millington, TN. The operational classification algorithm and the underlying rationale are described at length by EDS (EDS Federal, 2001). The purpose of the algorithm is to generate a ranking of the jobs to which a person should be assigned considering input personnel data. For every applicant $i$ and job $r$ the algorithm computes two utility functions $S_{ir}$ and $Q_{ir}$ based on the ASVAB and the AFQT scores, respectively. The $S_{ir}$ is designed to reward applicants with high ASVAB scores relevant to this particular job. The $Q_{ir}$ function is designed to discourage assigning overqualified applicants to easy jobs. The algorithm ranks the jobs for every applicant based on some combination of $S_{ir}$ and $Q_{ir}$. The basic algorithm has been modified to reduce the number of early-arriving highly capable persons that are assigned to jobs that are too “easy” for them. The modified algorithm, if a person is well qualified for a number of jobs, “pushes” him/her to the job where they are most “valuable” compared to other candidates.

Table 3 gives the results of the simulations.

\begin{table}[h]
\centering
\caption{Classification Simulation Results}
\end{table}
<table>
<thead>
<tr>
<th>Scenario #1: 1.7% less female jobs than females (8,134 jobs; 8,275 females)</th>
<th>Composite Set without AO or CS</th>
<th>Composite Set with AO</th>
<th>Composite Set with CS</th>
<th>Composite Set with AO and CS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassigned Recruits</td>
<td>469</td>
<td>413</td>
<td>389</td>
<td>288 - 303 (range with 4 runs)</td>
</tr>
<tr>
<td>Job Fill Standard Dev.</td>
<td>16.1%</td>
<td>15.1%</td>
<td>14.6%</td>
<td>13.7% - 14.8%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario #2: 2.5% more female jobs than females (8,484 jobs; 8,275 females)</th>
<th>Composite Set without AO or CS</th>
<th>Composite Set with AO</th>
<th>Composite Set with CS</th>
<th>Composite Set with AO and CS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassigned Recruits</td>
<td>501</td>
<td>440</td>
<td>279</td>
<td>279 - 300 (range with 4 runs)</td>
</tr>
<tr>
<td>Job Fill Standard Dev.</td>
<td>20.2%</td>
<td>18.7%</td>
<td>16.7%</td>
<td>16.4% - 17.4%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario #3: 6.2% more male jobs than males (38,402 jobs; 36,154 males)</th>
<th>Composite Set without AO or CS</th>
<th>Composite Set with AO</th>
<th>Composite Set with CS</th>
<th>Composite Set with AO and CS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassigned Recruits</td>
<td>938</td>
<td>661</td>
<td>785</td>
<td>492 - 555 (range with 4 runs)</td>
</tr>
<tr>
<td>Job Fill Standard Dev.</td>
<td>13.8%</td>
<td>13.4%</td>
<td>13.0%</td>
<td>12.6% - 14.4%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario #4: 13.4% more male jobs than males (40,995 jobs; 36,154 males)</th>
<th>Composite Set without AO or CS</th>
<th>Composite Set with AO</th>
<th>Composite Set with CS</th>
<th>Composite Set with AO and CS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassigned Recruits</td>
<td>387</td>
<td>71</td>
<td>213</td>
<td>0 (range with 4 runs)</td>
</tr>
<tr>
<td>Job Fill Standard Dev.</td>
<td>15.9%</td>
<td>15.6%</td>
<td>15.6%</td>
<td>18.2% - 19.3%</td>
</tr>
</tbody>
</table>

Table 3 gives a breakout of the number of unassigned recruits resulting from simulations that were performed using the Rating Identification Engine (RIDE) classification algorithm (Crookenden & Blanco, 2002). Gender classification was modeled separately (not presented here) because not all Navy jobs are open for women. The ratio of the number of jobs and the number of candidates were varied for the four runs to obtain a sense of the assignment outcomes for varied recruiting climates (enough and not enough recruits to fill Navy jobs). Four runs were conducted for Composite Set 4 to obtain a range of values that results from a new random seed initiated at the beginning of the classification simulation.

From Table 3 it is clear that in every classification simulation, providing an ASVAB composite set that included some composites with the AO or CS tests results in fewer individuals unassigned to jobs. Also, in general, the standard deviation of the fill of jobs (indicating evenness of distribution) tends to decrease with adding AO and CS tests. The obvious exception is for scenario #4 (13.4% more male jobs than males to assign) where everyone was assigned to a job using the AO and CS composite set, but with less of an even fill across ratings (standard deviations of from 18.2% to 19.4% for 4 runs, as compared to the high 15% for the other runs). Future research will examine this outcome in depth.

Conclusions

Including tests for classification purposes in the ASVAB, or as special tests administered along side ASVAB, improves the actual classification outcomes that are simulated for the Navy in terms of the number of individuals filling a set number of jobs. This outcome occurs when the
added tests lower the average intercorrelations of the existing battery. Additional benefits accrue if the tests lower barriers for minor groups to school entry and increase the validity of predicting the outcome of concern (for this study, Navy school performance and attrition). From this study, a recommendation could be made to review past military service tests that were not evaluated on all three criteria evaluated in this study.

References


The drop-out rate during training: in line with expectations?

C.P.H.W. van de Ven*

Introduction
Each year the Royal Netherlands Army, or RNLA, needs to take on approximately 3,000 personnel for the complement of privates/corporals with a fixed-term contract. This number will rise on the basis of the current Defence White Paper. To achieve this quota, a much larger group of interested people is needed in order to have enough candidates who maintain that interest during the application phase and to succeed in getting through the (demanding) physical and mental selection process. At the time that the candidate has completed the phase and is deemed trainable, he or she is of great value to the RNLA. This is due, on the one hand, to the investment already made in the potential fixed-term contractor and, on the other, to the need to fulfil the procurement requirement. The latter aspect is a particularly important factor in these times of a competitive labour market.

From the moment that the candidate is taken on, he or she attends a training course at one of the four school battalions of the Training Centre for Basic Training: Central, Airmobile, North or South. Following the three-month General Military Training, which is a few weeks longer at the Airmobile school battalion, the fixed-term contractors attend one or more functional training courses at the relevant training centres. Throughout the process, but chiefly during the General Military Training, personnel drop out of training. Among the 1998 intake group, the drop-out rate was on average below 15%. In 1999 this percentage is significantly higher, namely about 24%.

As holding onto personnel is increasingly the Achilles heel of the RNLA, the interim drop-out phenomenon is a problem for us all. This article devotes attention to the motives of fixed-term contractors for leaving training early, the increase in the drop-out rate during training and points for attention to reduce this drop-out rate.

* C.P.H.W. van de Ven (drs.) works as a researcher at the Behavioural Sciences Division/Directorate of Personnel and Organisation/RNLA.
Article structure

The article consists broadly of two parts: the theoretical part and the empirical part. The theoretical part comprises a description of research conducted in the past into the drop-out rate, resulting in a theoretical model. This theoretical model is then used as the basis for the empirical part. Here, the method of research is first looked at, followed by a description of the research results. The article ends with a few remarks and a conclusion.

Part I:
The drop-out rate in a theoretical perspective

For the last few years, a great deal of attention has been devoted to the drop-out rate among personnel. The emphasis in most studies is on personnel voluntarily leaving the organisation, whereby the employee takes the initiative to leave; this is in contrast to personnel being obliged to leave the organisation, whereby the initiative lies with the employer.

In the research under discussion, we are faced with a specific form of the drop-out rate, namely that which occurs shortly after entry into the organisation, during training. The employee has therefore not yet become acquainted with his or her eventual job or function. The theoretical description will therefore largely concentrate on this form of the drop-out phenomenon.

Drop-out models

A model which was advanced in 1973 by Porters and Steers assumes that employees have specific expectations on entering the organisation. If the organisation and the function do not meet initial expectations, dissatisfaction may arise, increasing the risk of drop out. Other researchers declare that dissatisfaction with work initially leads to the employee developing ideas about possibly leaving the organisation. These ideas lead in turn to estimating the expected value of looking for another job and the cost incurred by resignation. If these estimates are favourable, the employee will look for another job. In addition to the cost of resignation, what is to be gained by resignation and what is to be gained by staying on may also play a part in the estimate.
Any consideration of staying or leaving always includes the economy as a major factor. In times of high unemployment and few vacancies, fewer people will resign from jobs, even if they are dissatisfied. Under more favourable conditions, if there are many jobs available, dissatisfied employees will show a greater tendency to leave than satisfied employees. The relationship between satisfaction and the drop-out rate is therefore stronger in times of low unemployment than in times of relatively high unemployment. An additional theory is that there are different determining factors for dropping out in times of high unemployment than in times of plentiful employment. This means that, depending on the state of the economy, in this case the situation on the labour market, the weight applied to the determining factors may vary.

**Expectations**

A meta-analysis of 31 studies relating to the non-realisation of expectations resulted in the following findings: non-realised expectations have a strong influence on job satisfaction, involvement in the organisation, the intention to leave and the actual act of leaving the organisation. Incidentally, this only applies to expectations concerning aspects of the job which are deemed to be important. In the case of non-realised expectations of irrelevant aspects, the reaction is surprise rather than disappointment.

**Important aspects in a job**

In her study into “Important aspects in a job”, Van Zijderveld put the following question to young people in 1999: What are the three most important things you (would) look for when choosing a job?

<table>
<thead>
<tr>
<th>Top 3</th>
<th>Man</th>
<th>Woman</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Salary</td>
<td>88%</td>
<td>80%</td>
<td>84%</td>
</tr>
<tr>
<td>2. Atmosphere</td>
<td>52%</td>
<td>61%</td>
<td>57%</td>
</tr>
<tr>
<td>3. Interesting work</td>
<td>37%</td>
<td>43%</td>
<td>40%</td>
</tr>
</tbody>
</table>

Salary is spontaneously named most frequently, and more often by boys than girls. Girls seem to name this less frequently as they get older. For boys age makes no difference.
People with a technical education name salary relatively often in contrast to those with a care/healthcare education. Atmosphere (pleasant colleagues and working atmosphere) is named more often by girls than by boys. Interesting work (variation, interesting, fun and a challenge) is named about as often by both sexes. Young respondents name this more frequently than older respondents.

Van Zijderveld also reports on several other studies in which young people have been asked what they find important in a job. An initial measurement held in 1988 among students in their final year of further education highlighted the following job aspects, in order of importance.

1. a good income (in first place for 75% of the young people; boys seem to be more strongly interested in income than girls);
2. communication in the workplace;
3. involvement in content.

A study of the social environment of young people was conducted by means of interviews with 500 young people in 1988. The young people formed a random selection of Dutch young people between the ages of 12-21. The following ranking was drawn up on the basis of importance:

1. the nature of the work is pleasant, varied and fascinating;
2. good contacts between colleagues, good cooperation, pleasant atmosphere at work;
3. a good salary, earning money yourself, being financially independent;
4. a job must benefit well-being;
5. opportunities to develop;
6. bearing responsibility;
7. good working conditions.

Girls seem to have mainly intrinsic job expectations. These are expectations which relate to the nature and content of the job (such as variety in the work, good contacts with colleagues, independence and being appreciated). Boys have mainly extrinsic job expectations. These relate to job characteristics which serve as a means of achieving a
goal outside work (such as a good salary, working conditions and having a career). Their own income is deemed important by more boys than girls.

With respect to what young people think important in their jobs, social milieu forms the basis for a clear difference:

a. Young people from higher social milieus are more development and career-oriented than those from lower social milieus. The higher the social milieu, the more frequently young people emphasise communication, involvement, management, status and career.

b. Middle-class young people think that having a career is more important than those from a higher class, and they want to climb the social ladder via their careers.

c. Young people from the lower class are more materiel-oriented and place a good income at the top of their list.

Young people who start work at a young age (these are often lower class) often have a fairly instrumental attitude to work, while those who start later are more interested in content.

Important issues concerning the future, such as choosing a partner, may also influence what people think important in a job.

In a longitudinal study held in 1988 in the school battalions, Roepers and Duel investigated which characteristics of a job are seen as the most important. The table below gives these aspects and the point at which the fixed-term contractors were questioned.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Most important aspects of a job at different points</th>
</tr>
</thead>
<tbody>
<tr>
<td>week 1 (first day)</td>
<td>week 5</td>
</tr>
<tr>
<td>1. teamwork</td>
<td>1. salary</td>
</tr>
<tr>
<td>2. camaraderie</td>
<td>2. camaraderie</td>
</tr>
<tr>
<td>3/4. gaining work experience</td>
<td>3. variety</td>
</tr>
<tr>
<td>3/4. being able to become an</td>
<td>4/5. getting the job you want</td>
</tr>
</tbody>
</table>
The importance that one attaches to specific job characteristics is clearly related to the point, or the state of progress, at which they are questioned in the training process. In the first week, aspects related chiefly to the working atmosphere are found the most important (teamwork and camaraderie). A notable absence in the top 5 at this point is salary. From a certain point, which cannot be specified more closely using these data, the importance of salary increases (enormously) and, for the fixed-term contractors in training, this becomes the most important aspect of their job. Camaraderie is deemed the second most important at all three measurement points.

**Psychological contract**

In a study previously carried out into the drop-out phenomenon among the school battalion South, in which salary is often cited as a reason for leaving, the importance of the salary is linked to the employees’ psychological contract.

The psychological contract comprises the unspoken (often unconscious) expectations with respect to mutual obligations held by the organisation and the individual. The individual commits to making certain sacrifices and, in return, obliges the organisation to guarantee a good salary, social security and security on the duration of the work. In psychological contracts a distinction is made between transactional contracts and relational contracts. Transactional contracts are chiefly economic and extrinsically oriented and are specified in time and time-related. This type of psychological contract is mainly found among employees who are taken on by the organisation for a specified period with a view to performing special tasks, for which specific expertise and skills are required. These employees are prepared to work hard in exchange for significant financial rewards. Relational contracts are both economically and non-economically-oriented, but otherwise have a more socio-emotional and intrinsic nature and are open-ended (in the case of contracts for an indefinite period).
It can therefore be assumed that young people in particular who are in service for a short period will develop a more transactional contract. An important expectation of this category of employees is that they are awarded reasonable financial compensation for their efforts. If this does not happen, they will either (try to) terminate the working relationship or reduce their effort to the level at which the employee believes the contract is once more properly balanced.

Two thirds of Dutch young people indeed estimate that fixed-term contractors earn a good salary. This has been proved by studies into the interest of Dutch young people in a job, conducted in the context of the Defence Labour Market Monitor 1999. About a quarter of the respondents had no clear picture of this and only a very small percentage thinks that fixed-term contractors have no good salary.

**Orientation**

Earlier research into the drop-out phenomenon in the school battalions shows that in particular the first week, and to a lesser extent the second week, is prone to the drop-out phenomenon. The initial period and in particular the first week therefore require additional attention.

The first week of entry into an organisation is known as the orientation period in the literature on the subject. This period is treated as a separate entity as this week goes hand-in-hand with a high degree of stress. This stress is caused by the large number of role changes experienced by the individual. Changes which are known to possess a specific stress value and could perhaps occur in new recruits include: the first full-time job or new job, geographical move, earning an income or an increase or drop in income, serious limitation of social life and separation from the parental environment. Furthermore, stress occurs due to the non-realisation of expectations and as a result of worries that one is incapable of meeting the requirements of the organisation.

**Individual factors**

A number of demographic and person-related variables may play a part in people leaving the organisation. Firstly, research has shown that the drop-out rate is higher among
younger employees than among older employees. This relationship may partly be explained by other variables. A younger employee generally has more opportunities and fewer responsibilities than an older employee.

There is also a negative link between period of service and leaving the organisation. The drop-out rate is relatively higher in the early years of service than in later years. Demographic variables such as sex, education and civil status may also have an effect. No systematic links have been demonstrated between personality traits, interests, intelligence and capacities on the one hand and the drop-out phenomenon on the other. This has also been proved by previous research by the Behavioural Sciences Division.

Information

The adaptation of newcomers during the initial period depends partly on the precision and completeness of the information given to the individual before entering service. Incomplete and imprecise information can lead to non-realisation of expectations and a reality shock. The extent to which the newcomer has realistic, precise and full information at the time of entering service results in positive attitudes and a lower drop-out rate.

Job versus organisation

Earlier research into the drop-out phenomenon highlighted the fact that some of those who decide to leave the RNLA do so on the basis of not obtaining the function they wanted.

The choice of a specific job in a specific organisation is the result of a series of decisions made during the period of growth into adulthood: this is known as the exclusion process. During childhood, all or nearly all professions are a possibility for the ultimate choice. During puberty, this choice is reduced to the choice for a general professional sector, and in the young adult stage the choice of a specific profession is made. During adulthood, this choice is refined further into the choice of a specific job and the related organisation. The essence is that people vary with respect to the image they have of their work. The most general image relates to a desired professional sector (military). A further refinement applies if people have already opted for a specific profession (military driver).
Finally, people may already have made a choice of a specific job and the related organisation (military truck driver).

*Research model*

The principle of the theoretical description is that the extent to which the initial expectations can be met determines how satisfied people will be with their new job and ultimately may determine their decision to leave the organisation. This only applies to the expectations concerning job characteristics which are deemed important.

Salary is still viewed as the most important or one of the most important characteristics. The importance of salary increases the more people’s profiles meet the characteristics given below:

1. young people who come on the labour market at an early age (often the lowest qualified);
2. male;
3. young people from a lower social milieu;
4. people with a contract for a specific period.

In the case of a job as a fixed-term contractor, the fact that Dutch young people expect fixed-term contractors to earn a good salary is also added. This is a positive aspect for recruitment, but this may increase the drop-out rate in view of expectations on this aspect possibly being too high and therefore being impossible to realise.

Other aspects which young people in general and fixed-term contractors in particular deem important are a good working atmosphere (teamwork and camaraderie) and pleasant, interesting work (getting the function you want).

On the basis of the ranking according to importance of job characteristics (table 2), it can be expected that at the start of training reasons for leaving will chiefly be named which relate to the working atmosphere. Although disappointing experiences with respect to camaraderie are a possible reason to leave throughout training, (possible) disappointment...
about salary will have an even greater impact, with the exception of the period at the start of the appointment.
Explanation of model (from right to left)

People do not leave training or the organisation simply for one reason or for clear reasons. People who leave training or the RNLA due to a reason which lies outside their sphere of influence or outside the RNLA do not want to leave the organisation due to dissatisfaction (in the model: satisfaction). Yet for these people, too, something has happened which is not in line with their expectations (in the model: realised expectations). In addition to the various reasons which may be the cause of resignation, the reason for leaving may also be (strongly) related to the point in time of resignation. During the orientation period, other reasons may also be decisive than those which apply in the period which follows.

Only non-realised expectations relating to important aspects of the job (salary, working atmosphere and nature of the job) will lead to dissatisfaction and possibly ultimately to leaving the organisation.
A (more) competitive labour market will result in an increased tendency for people to leave the organisation. This means that static job satisfaction in times of an increasingly competitive labour market leads to a higher drop-out rate.

The extent to which expectations can be realised goes hand-in-hand with how realistic these are. Expectations which are too high are difficult to realise and therefore lead to disappointment more quickly. Expectations can be aroused by recruitment and information, correspondingly unrealistic reasons to apply for other jobs, and individual factors.

Part II:
The empirical study

Study method
What are the motives for fixed-term contractors to terminate their training early? Can the increase in the drop-out rate be explained? What are the areas for attention to reduce the drop-out rate (in the short term)?
In order to be able to answer this question, the study model given earlier was developed on the basis of the theoretical description. The basic idea of the model is:

The drop-out rate, during the initial period of appointment, is determined by the extent to which initial expectations concerning relevant job aspects are realised.

This applies both if the initiative comes from the individual and if the organisation dismisses the person. Both parties, in principle, have a positive view of each other prior to the act of appointment. If the individual subsequently takes the initiative to leave, then there is something or several aspects which appear not to have gone as he or she had expected, otherwise the person would have stayed on. If the organisation dismisses a
person, the organisation therefore experiences a discrepancy between the requirements and the skills of the person, i.e. the person does not meet expectations. According to the above, there are just as many reasons for leaving as there are relevant job characteristics, or combinations of job characteristics. In order to generate usable data, it is therefore important to discover whether groups of individuals can be distinguished which experience comparable discrepancies.

**The instrument for data collection**

Using the study model and the job characteristics earmarked as relevant, a questionnaire was drawn up in which these aspects have been included. These exit questionnaires are given to all fixed-term contractors who leave training early by employees of the Total Quality Management Office of the relevant school battalions. The response percentage is therefore 100%.

**Those who continue**

In addition to those who leave training, a representative random selection is made from those who successfully complete training and they are also asked to complete the questionnaire. This structure was chosen in order to be able to relate and put into perspective the answers of those leaving training. This structure and the continuous nature of the study mean that the data are also highly suited to evaluating training (continuously).

**Timescale**

The timescale in which the data for this study were collected comprises the period from March 1999 up to and including February 2000. In total, the data from 601 completed questionnaires by those dropping out of training and 389 of those continuing were analysed*.

---

* The result section of this study was created by analysis of data input up to and including February 2000. Data for the whole of 2000 will soon be available for analysis. It should therefore be noted that measures already taken over the course of 2000 may result in a few alterations to the situation.
**Study results**

Fixed-term contractors do not appear to leave training early due to one reason or for clear reasons. In most cases, it is a combination of reasons, in which different motives play a part for different people. On average, a student has 2.6 reasons for leaving training early. The most frequently cited are ‘military life does not suit me’(46% of all those who left named this as one of their reasons), ‘homesickness’ (28%), ‘circumstances at home’ (24%) and ‘disappointing salary’ (24%).

A few clear differences can be distinguished between men and women. Women more often cite: ‘medical reasons’ (27% versus 14%) and ‘physical load’ (29% versus 12%). In contrast, men more frequently cite ‘disappointing salary’ (26% versus 7%).

The fact that fixed-term contractors have several reasons for leaving training creates the need to investigate whether there are common combinations of reasons for leaving, in order to provide insight and in the interest of practicality. The cluster analysis shows that those leaving training can indeed be classified into five categories or clusters. The following reasons for leaving are not included in the analysis: ‘different function within the RNLA’, ‘dismissed’, ‘temporary stop’ and the reason ‘other’. For the first three reasons, this is not a (definitive) loss for the organisation. The reason ‘other’ is too diverse to use as a variable in a statistical analysis.
Table 3  Clusters of reasons for leaving

<table>
<thead>
<tr>
<th>Cluster 1 (n = 127)</th>
<th>Cluster 2 (n = 122)</th>
<th>Cluster 3 (n = 88)</th>
<th>Cluster 4 (n = 59)</th>
<th>Cluster 5 (n = 90)</th>
</tr>
</thead>
<tbody>
<tr>
<td>homesickness</td>
<td>military life does not suit</td>
<td>circumstances at home</td>
<td>wrong function</td>
<td>disappointing salary</td>
</tr>
<tr>
<td>possible mission abroad</td>
<td>medical reason</td>
<td>homesickness</td>
<td>few career opportunities</td>
<td>military life does not suit</td>
</tr>
<tr>
<td>military life does not suit</td>
<td>physical load</td>
<td>possible mission abroad</td>
<td>starting studies</td>
<td>better job</td>
</tr>
<tr>
<td>posting to Germany</td>
<td>better job</td>
<td></td>
<td>few indefinite-term contract opportunities</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>few career opportunities</td>
<td>starting studies</td>
</tr>
</tbody>
</table>

Each of the clusters is described below according to their characteristics.
Next, there is a discussion of the extent to which the entire study population’s expectations were realised with respect to the most important job aspects, as well the extent to which this applies to the various clusters.

Cluster 1  Homesickness/mission abroad

Quotation from training drop-out: “I am suffering from homesickness because the situation is good at home and not here.”

and

“I miss my home. I realise that I cannot cope with the six-month mission abroad. I do think that the RNLA is a good employer though”.

In the first cluster, homesickness plays a central part. Of the 127 people in this group, 113 name this as a reason for leaving. Homesickness is often combined with the reason of the ‘possible mission abroad’ and/or ‘military life does not suit me’.

Of the entire exit group, almost one third decides to leave training as early as in the first two weeks. Of this group, 40% come from this cluster. Within cluster 1 a distinction can also be made between two sub-clusters. The first does not mention homesickness in combination with a possible mission abroad, and those in the second sub-cluster do, as may be expected, name homesickness in combination with a possible mission abroad.

The two clusters differ significantly with respect to the point at which they leave: of those who mention only homesickness (or in combination with another reason for leaving), two
thirds decide to leave in the first two weeks, in comparison to one third of those who also mention ‘possible mission abroad’.

In addition to the finding that people who leave training early are significantly younger than those who successfully complete training, those in cluster 1 are predominantly quite young. The average age is 18½ years, in comparison to the average of 19 years of the total exit population and 20 years of those who continue (the latter are also, however, already a maximum of six months longer in service). As can be expected, the share of 17-year olds in this cluster is also greater than in the other clusters. Related to the above, cluster 1 contains a relatively high number of single young people and those still living with their parents.

In addition to the finding that those in cluster 1 are less positive about working hours, this indicates that the expectations of a large number of people in this cluster did not come true, or not fully, with respect to the working atmosphere. Those in cluster 1 particularly feel, more than those in other clusters, that they were awarded too little respect and appreciation and that there were too few opportunities for showing initiative (in a broader sense: being able to take decisions themselves and not be ordered around all the time). Of all the clusters, those people in cluster 1 experienced their period in service the least positively.

**Cluster 2 Military/physical**

Quotation from training drop-out: “You have a totally different picture of this work and I realised that I cannot cope physically”.

The most commonly named reason in cluster 2 is: ‘military life does not suit me’. This reason is chiefly combined with medical reasons and/or physical load. Proportionally speaking, it is often indicated in this cluster that the person involved has already found a better (or physically less demanding?) job.

Although 30% of those from cluster 2 had already taken the decision in the first two weeks, a relatively high percentage actually leaves in weeks 3, 8 and 9. These are the times during the General Military Training, or immediately after, that bivouacs take place and the physical load is increased.
Within the entire group of those who leave training, there are relatively more women than in the group of those who continue. Percentage-wise, most female training drop-outs occur in this cluster. Viewed as a whole, the people in this cluster are satisfied with the various aspects of the job. They do, however, indicate having significantly more difficulty with the pace of the General Military Training, physical education and sport and field service/exercises.

**Cluster 3  Circumstances at home**

Quotation from training drop-out: “I’m having problems at home with my girlfriend, she cannot cope with me being away from home all week”.

The people within cluster 3 all indicate that they left training due to circumstances at home. This often goes hand-in-hand with feelings of homesickness and (problems with) a possible mission abroad.

Although it may be expected that the point of departure for persons in cluster 3 is evenly spread over training, a relatively higher number of people decide in weeks 5 and 6. The average age of cluster 3 is the highest of all the clusters and comparable to that of those who continue. With respect to previous education, this cluster is characterised by a relatively large number of people with only lower education, as well as a relatively high number with an intermediate vocational education diploma. A relatively high number have already left the parental home. This cluster is also characterised by a high degree of work experience. Various work-related aspects are appreciated the most by people in this cluster, and they are also the most positive about the period in service. In contrast, they can rely the least on support from home with respect to the choice of a job as fixed-term contractor and the possibility of being sent on a mission abroad.

**Cluster 4  Did not get desired function**

Quotation from training drop-out: “A great deal of fuss about functions which are available, but which are allocated with taking the persons themselves into account”.

All those within cluster 4 give as a reason for leaving that they did not obtain the function they had expected or which they had opted for. The reasons of ‘few career opportunities’,
‘starting studies’ and ‘few opportunities to become indefinite contractor’ are also relatively common in this cluster.

No-one in this cluster took the decision to leave as early as the first week. The weeks in which people dropped out most were the weeks (7 to 9) in which function allocation was carried out. There are relatively few women in this cluster. The previous education of those in this cluster is typical. Within cluster 4 there are relatively few people with a general education (lower/higher general secondary education) and almost half joined up immediately after leaving school, compared to 29% in the other clusters. In comparison to the other clusters, appreciation of working conditions, with the exception of working hours, is lower. This cluster naturally scores poorly in the question of whether they had obtained the desired function.

Cluster 5 Disappointing salary
Quotation from training drop-out: “The salary is far too low for the work you have to do here”.

‘Disappointing salary’ is named by everyone in cluster 5 as a reason for leaving early. The disappointing salary is often combined with ‘military life does not suit me’ (or is that part of military life?). The reasons of ‘better job’, ‘few career opportunities’ and ‘starting studies’ are mentioned relatively frequently in this cluster.

The drop-out phenomenon in cluster 5 manifests itself in two periods. The first period is week 2, the point at which the salary official calculates how much the fixed-term contractors will earn after taxes etc., and the second is weeks 5 up to and including 7, the period in which the first bank statements reach the fixed-term contractors.

With respect to personal traits, this cluster is distinguished by the very small number of women, the relatively high age and the correspondingly small percentage of 17-year olds. A relatively high number of them have had some other work experience prior to this. It is clear that in particular in this cluster, expectations of working conditions come true to a lesser extent: 74% indicate that these expectations do not, or scarcely, come true. The same percentage indicates dissatisfaction with the salary and the average score is 4.1.

Almost everyone in this cluster thinks that they can earn more outside the RNLA. With respect to the working atmosphere and the working relationships, a large number indicate
that they are dissatisfied with the (respect and) appreciation awarded them. As far as training is concerned, a relatively high percentage say that they had difficulties with the study and evening workload.

**Important job aspects**

The most important job aspect about which the target group has expectations is a **good salary**. This is, however, also the aspect about which there is the greatest disappointment. Of those leaving training and those who continue, almost half think that expectations concerning the working conditions do not, or scarcely, come true. Women are, incidentally, significantly more positive about the extent to which this expectation is fulfilled (26% versus 49%).

With respect to appreciation of the salary, those who complete training are more negative than those who left training early (4.6 versus 5.6). Account needs to be taken here of the fact that about half of those who leave do so before the first salary is paid.

To the question of where fixed-term contractors expect to earn more, almost threequarters say outside the RNLA, only 5% think that they could earn more within the RNLA.

With regard to a **good income**, it is clear that all those in cluster 5 (disappointing salary) are highly disappointed and, as it were, forced out of the organisation. Disappointing salary is in this case a *push factor*. In view of the low appreciation of the entire group (those who leave and those who continue), but mainly the fact that a majority thinks it can earn more outside the RNLA, the more attractive salaries elsewhere act as a *pull factor*. It must not be forgotten here that it is precisely salary which is the most important form of reward for this group.

The second most important job aspect is a **pleasant atmosphere in the workplace**. Here there is a difference of opinion between those who leave and those who continue. One in three of those who leave early indicates that expectations of atmosphere did not come true, or not fully, compared to 14% of those who continue. The appreciation awarded is unsatisfactory to a relatively large number of people. Incidentally, the other aspects relating to the working atmosphere, such as colleagues and immediate superiors, were awarded high scores by both those who left and those who continued.
Non-realised expectations concerning the **working atmosphere** are an important factor in leaving training for those in cluster 1 in particular (*homesickness/possible mission abroad*). Many experienced the strict military life as such a huge shock on entering service that they decided to leave as early as in the first week, or in the second week. The phenomenon of homesickness can generally be sub-divided into two factors: difficulty in being away from the old, familiar (often parental) environment and difficulty in adapting to a new environment. In view of the low appreciation during this period and the non-realisation of expectations surrounding the atmosphere, the difficulty in adapting to the new situation is more crucial here. Furthermore, the more pleasant the old situation was and the more unpleasant the new situation is for these people, the more difficult it is to make the transition.

As stated earlier, a distinction can be made for cluster 1 between those with homesickness and *no* combination with a possible mission abroad and those *with* the combination of the two. Of those who only named homesickness as a reason for leaving, 67% leave within two weeks. They do not succeed in adapting.

Those with homesickness in combination with a possible mission abroad have succeeded to some extent in adapting to the new environment; they leave at a later point in training. The anticipatory idea of being sent on a mission abroad to a (possibly) stressful environment, whereby their separation from the familiar environment is irrevocable, makes them decide to leave training and the RNLA.

The third job aspect which is very important to fixed-term contractors is the **content side of the work**: it should be fun, varied and interesting. Although the training period is different from the combat-ready period, it is important here to investigate the extent to which there is possible dissatisfaction among the study population. Here a distinction has been made between work content and work relationship.

With respect to the work *relationship*, there are two main aspects which require improvements. The first concerns the respect and appreciation awarded. A considerable percentage of the exit group indicates an insufficient degree of respect and appreciation. Secondly, a large number believe that there are too few opportunities for showing initiative (and more broadly also to take decisions themselves and not be ordered around
all the time). The most important aspect, being able to obtain assistance from NCOs/officers, is appreciated positively.

With respect to the important aspects relating to work content, almost everyone agrees that this is at least sufficient.

There is a strong relationship between job content and whether people are given the function they opted for. Among those who left training early, the percentage which indicates that it did not obtain the desired function is twice as high as among those who continue.

For cluster 4 (not getting desired function) the motive for resigning is clearly taken from disappointment about (future) job content. Military life and the RNLA as an employer appeal greatly, but the future function they have been allocated does not correspond with their expectations and is sometimes even viewed as a broken promise.

For cluster 2 (military/physical), too, there is a discrepancy between the expected job content and the reality. This is, however, different from that in cluster 4 with respect to the future function. Those in cluster 2 have difficulty with the current job content.

Aspects of training are either too difficult or the pace is too high.

**Other work-related aspects**

Those in cluster 3 (circumstances at home) also have a number of actual expectations which did not come true or a number of incidents occurred which were not in line with expectations. In most cases, however, these were beyond the sphere of influence of the RNLA or of the person involved. These often concerned deaths, divorces, illness in the immediate family or the partner not agreeing with the choice of a job as fixed-term contractor, with the corresponding possibility of being sent on a mission abroad for a long period. All these reasons are directly or indirectly related to the RNLA, but in particular to the characteristics of a fixed-term contract job: a 24-hour a day job which is sometimes difficult to combine with a family life.

**Training**

Although the general assessment of General Military Training and Functional Training is satisfactory, there are number of aspects which may aid the drop-out rate. For instance,
more than a quarter of those leaving training indicate that they found the pace of General Military Training too high, compared to 5% of those who continue. The study and evening workload is experienced as considerably higher by those who leave than by those successfully complete training. With respect to the physical load during General Military Training, those who drop out indicate more frequently that the aspects in question are ‘too difficult’, or ‘too fast’. The greatest difference concerns ‘field service/exercises’: 24% of those who leave training find this too difficult compared to 3% of those who continue. The acceleration in training is viewed as ‘too fast’ by 21% of those who leave compared to 10% among those who continue.

Students are confronted with a busy schedule from the beginning. Although it is clear that it would be desirable for several reasons to start more slowly, especially in the early stages, in view of the content of General Military Training and the ultimate goal, there is simply not enough time for this.

The consequences of this strict schedule are that an increased number of people leave due to physical complaints (including many female fixed-term contractors) and too high a mental load. In addition, this busy schedule offers no room for acclimatisation (reality shock) to the new living and working environment. In particular during the first week (known as the orientation phase in the literature on the subject) activities should be aimed at assisting the newcomer in conquering the stressful new situation. But after the first week, too, the lack of time means that there are too few times of rest and free evenings in which a pleasant working atmosphere can develop.

**Increase in the drop-out rate**

If the labour market is becoming increasingly competitive, static job satisfaction leads to a higher drop-out rate. This also applies to the RNLA, even if nothing has changed in comparison to 1998, and an increase can be explained by the large(r) number of vacancies elsewhere. It is clear that the economy is a difficult factor to influence. In order to even neutralise the drop-out rate due to this reason, job satisfaction needs to be increased.
Notes
The study results clearly show that life and work in the military organisation has many positive aspects and is appreciated as such. Even within the group of those who leave training, a large number are highly positive about their period in service. Work content aspects and collegiality are particularly experienced as positive by almost everyone. The aim of this study, however, is to highlight the problems. This inherently means that little attention is devoted to positive aspects, but these are clearly present during training.

Although the expectation is that the drop-out rate can be reduced with a number of (radical) measures, it is not realistic to assume that it can be brought down near to zero. There are too many factors of influence on the drop-out rate which cannot, or only to some extent, be influenced, including home-related aspects.

It must be remembered for all the measures that problems experienced by one person may be normal for another. When taking measures, it is therefore important to ensure that the solution for one problem does not create any problems for another.

To recap
The extent to which initial expectations concerning relevant job characteristics are fulfilled determines to a great extent the (job) satisfaction and the corresponding drop-out rate. This relationship is stronger the more jobs are available on the job market. This means that static job satisfaction in the event of an increasingly competitive labour market results in the drop-out phenomenon. The non-realisation of these expectations and the resulting reasons for leaving are based on a number of bottlenecks. Removing these bottlenecks can raise (job) satisfaction and thus reduce the drop-out rate among fixed-term contractors in training.

The areas for attention are: working conditions, these lead to dissatisfaction and cannot compete with salaries in civilian society; function allocation, this should be carried out more in conjunction with the fixed-term contractors themselves, and training, the ultimate goals and the time reserved to achieve them do not correspond well. The pressure of time which is thus created leads to a number of undesirable developments, including: a reality shock at the start of training, too high and fast an increase in the
physical and mental load and too few rest periods to allow a pleasant working atmosphere to develop.

On the basis of these and other study findings, the RNLA has already taken a number of measures, and others still need to be implemented. A follow-up study will soon prove whether, and to what extent, the measures already implemented have had an effect.
ASSESSMENT OF THE POTENTIAL VOLUME OF RELEASES
AT THE 20 YEARS OF SERVICE POINT

Irene A. Collin and Paul Bender
Directorate of Strategic Human Resources
National Defence Headquarters, Ottawa, Ontario, Canada

ABSTRACT

The Force Reduction Program (FRP), which encouraged retirement and limited recruiting, affected various military occupations from 1992 to 1996. This resulted in less than ideal demographic profiles for the affected occupations. Currently, there is a concentration of Canadian Forces (CF) members between 10 and 20 years of service (YOS), and an unusually low number of members below 10 YOS. This paper attempts to quantify releases over the next 10 years, per occupational group, as CF members reach the 20-YOS point.

It is found that, not only is there an exceptionally large group approaching the 20-YOS point, but recent attrition rates at this point are higher than past attrition rates. During the next five years, overall releases among officers are expected to increase by 36%, and releases among non-commissioned members (NCMs) are expected to be more than double past release numbers.

I. INTRODUCTION

Background

1. The Force Reduction Program (FRP), which encouraged retirement and limited recruiting, affected various military occupations from 1992 to 1996. This and subsequent reduction measures resulted in the Canadian Forces (CF), which had nearly numbered 90,000 members, dropping to below 60,000 members. Another result has been less than ideal demographic profiles for the affected occupations, and currently, there is a concentration of CF members between 10 and 20 years of service (YOS), and an unusually low number of members below 10 YOS. Indeed, the proportion of the CF population currently between 15 and 20 YOS is the highest it has been since the late 1970's.

2. Various factors, such as the high rate of personnel tempo and the progressive erosion of the military way of life, may contribute to increased rates of attrition at all stages of the military career. However, the greatest impact of such attrition is expected to coincide with the highest concentration of members (i.e., 10-20 YOS). Historically, attrition rates drop steadily after 10 YOS and rise sharply at 19 YOS. Therefore, assessing attrition at the 20-YOS point should provide valuable information, in particular for long-term recruitment requirements.
Aim
3. The aim of this paper is to quantify anticipated releases over the next 10 years in order to assess the impact of the high concentration of CF members approaching the 20-YOS point.

Scope
4. This study includes all forms of attrition, whether voluntary or scheduled, occurring between 19 and 23 YOS. All military occupations are considered, divided into officers and non-commissioned members (NCMs), and into officer and NCM occupational groups. Occupational transfers are not considered as part of this study.

II. METHODOLOGY

Attrition Rates
5. The first step in attempting to quantify future releases was to determine appropriate attrition rates. To do this, total releases and total strength were extracted from the Directorate of Strategic Human Resources (D Strat HR) historical database, and attrition rates were calculated (attrition rate = releases/strength). These calculations were done for each occupational group and for the entire NCM and officer population, per year, from 1982 to 2000. (Data from 2001 were incomplete at the time of this calculation and therefore, unusable.) Annual attrition rates from 19 to 23 YOS, for all NCM occupations and all officer occupations, are shown in Figure 1.

![Figure 1: Historical Attrition Rates (19-23 YOS)](image-url)
6. It is obvious from Figure 1 that attrition rates in recent, post-FRP years have been higher than attrition rates prior to the FRP. Consequently, it was decided to use recent attrition rates in future projections. Releases and strength were again extracted from the D Strat HR historical database, for each occupational group and for all NCMs and officers, for each year of service from 10 to 23 YOS (in order to track the effect of releases on the population from 2002 to 2011), for years 1998 to 2000. Again, attrition rates were calculated as the ratio of releases to strength.

Initial Population

7. The initial population was extracted from the PeopleSoft database, as of January 2002. The population was distributed by YOS, by occupational group and by NCM or officer. Only the population within the range of 10 to 23 YOS was required to perform the 10-year projections.

Computation

8. The 2002 population matrix was multiplied by the attrition rate matrix to estimate the 2002 releases. Releases between 19 and 23 YOS were totalled and saved. The 2002 release estimate was subtracted from the 2002 population, and all entries were incremented by 1 year of service to estimate the 2003 population. This procedure was repeated until the 2011 releases were estimated. Forecast 19-23 YOS releases were then compared to past 19-23 YOS releases. For this comparison, releases from 1988 to 1991 were used since, before 1988, some members were on “old terms of service (TOS)” contracts, and after 1991, the FRP was in effect.

III. RESULTS

Caveats

9. The results of this study were generated by applying an average of recent attrition rates to an extraction from the PeopleSoft database (January 2002). They are, therefore, only predictions based on past trends. At any time, certain factors, such as alternate employment opportunities, perceptions of the military quality of life, and TOS conversion rates, may cause these patterns to change.

Release Projections

10. Figure 2 is a graph of historical releases for the period of 1982 to 2001, and forecast releases for 2002 to 2011, all between 19 and 23 YOS. From casual observation, the forecast releases appear to be roughly double those of the past. The greatest release estimates appear to occur in 2002, slowly tapering off to a more normal level in 2011.
11. Figure 3 presents a similar graph of officer releases. Forecast releases appear to be greater than past releases, but not alarmingly so.
12. Among NCM and officer occupational groups, the group that seems to present the greatest problem is Land – Combat Arms. Figures 4 and 5 show release profiles for the Land – Combat Arms NCM and officer groups, respectively. In both of these graphs, but in particular the NCM graph, forecast releases are expected to be much greater than past releases. At least in the case of the officer group, releases are expected to diminish in the future. However, it is not apparent whether the NCM release numbers are climbing, remaining steady at an unprecedented level, or eventually returning to historical levels.

![Releases, Actual and Forecast, Land - Combat Arms (NCM)](chart)

Figure 4: Land – Combat Arms (NCM) Releases, Historical and Forecast
13. In order to determine how serious a problem these potential releases might be, if a problem at all, a rating system evolved. It was decided that if release projections were within 10% of past releases, there was no problem to consider (coded green). If release projections were between 10% and 50% of past releases, there might be some concern (coded yellow). If release projections exceeded past releases by more than 50%, greater attention might be required (coded red). There are two exceptions to these rules, to account for small numbers. First, if the difference between projected and past release numbers was less than 10, the rating was downgraded by one level. Second, if the difference between projected and past release numbers was less than 5, the rating became green, regardless of the comparison between the numbers.

Analysis

14. NCM release projections are compared to past releases in Table I. As previously stated, past releases are average annual releases from the period of 1988 to 1991. The release projections are divided into two five-year periods, 2002 to 2006 and 2007 to 2011. The ratio of projected to past releases is computed and colour-coded. The proportion of the population between 15 and 20 YOS is also listed to provide some conception for the magnitude of these projections.
TABLE I
NCM RESULTS

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Air-General</td>
<td>28.18%</td>
<td>5.3</td>
<td>14.0</td>
<td>2.67</td>
<td>9.2</td>
<td>1.75</td>
</tr>
<tr>
<td>Air-Technical</td>
<td>36.80%</td>
<td>63.0</td>
<td>179.8</td>
<td>2.85</td>
<td>105.2</td>
<td>1.67</td>
</tr>
<tr>
<td>Elec./Comm.</td>
<td>32.10%</td>
<td>52.8</td>
<td>109.0</td>
<td>2.07</td>
<td>96.0</td>
<td>1.82</td>
</tr>
<tr>
<td>Engineering</td>
<td>29.78%</td>
<td>33.8</td>
<td>43.6</td>
<td>1.29</td>
<td>21.8</td>
<td>0.65</td>
</tr>
<tr>
<td>Flight Crew</td>
<td>36.82%</td>
<td>11.5</td>
<td>19.0</td>
<td>1.65</td>
<td>9.4</td>
<td>0.82</td>
</tr>
<tr>
<td>Land-Combat Arms</td>
<td>20.59%</td>
<td>32.5</td>
<td>139.8</td>
<td>4.30</td>
<td>158.4</td>
<td>4.87</td>
</tr>
<tr>
<td>Land-Maintenance</td>
<td>25.54%</td>
<td>34.5</td>
<td>71.6</td>
<td>2.08</td>
<td>71.0</td>
<td>2.06</td>
</tr>
<tr>
<td>Logistics</td>
<td>34.41%</td>
<td>140.3</td>
<td>275.4</td>
<td>1.96</td>
<td>199.0</td>
<td>1.42</td>
</tr>
<tr>
<td>Medical</td>
<td>28.27%</td>
<td>28.3</td>
<td>44.8</td>
<td>1.59</td>
<td>38.4</td>
<td>1.36</td>
</tr>
<tr>
<td>Sea-Combat/Gen.</td>
<td>23.73%</td>
<td>13.3</td>
<td>53.2</td>
<td>4.02</td>
<td>58.4</td>
<td>4.41</td>
</tr>
<tr>
<td>Security/Intelligence</td>
<td>32.51%</td>
<td>21.3</td>
<td>31.6</td>
<td>1.49</td>
<td>26.8</td>
<td>1.26</td>
</tr>
<tr>
<td>Specialist</td>
<td>28.29%</td>
<td>0.8</td>
<td>3.6</td>
<td>4.80</td>
<td>1.4</td>
<td>1.87</td>
</tr>
<tr>
<td>Sea-Technical</td>
<td>24.66%</td>
<td>18.5</td>
<td>47.8</td>
<td>2.58</td>
<td>54.0</td>
<td>2.92</td>
</tr>
<tr>
<td>All</td>
<td>28.83%</td>
<td>455.5</td>
<td>1033.2</td>
<td>2.27</td>
<td>849.0</td>
<td>1.86</td>
</tr>
</tbody>
</table>

15. Table I shows that, over the next five years, NCM releases are expected to be more than double those of the past. Out of the 13 occupational groups, eight are expected to be coded red. The other groups are coded yellow with the exception of the Specialist group, which would be “red” on the basis of its ratio, but is “green” by virtue of its small numbers. Indeed, two of the four “yellow” groups would be “red” but for the fact that small numbers downgrade their status to “yellow”. All groups are expected to improve over the second five-year period, with the exception of Land – Combat Arms, Sea – Combat Arms/General and Sea – Technical. The group of greatest concern would seem to be Land – Combat Arms, in consideration of its overall numbers and high release ratio, as future releases are expected to be more than four times those of the past.

16. Table II provides the officer version of these results. The dominance of the “green” occupational groups attests to the fact that most officer occupations have not been allowed to depart so drastically from the ideal YOS profile as most NCM occupations. Also, officer occupations are generally smaller in terms of manning level. Out of the 18 occupational groups, there are actually 11 with ratios in the red zone. However, so few releases are involved that only one of these is coded red: Land – Combat Arms. This explains why the overall officer rating is “yellow”, whereas most of the officer groups are coded green. Practically all occupational groups improve during the 2007 to 2011 period, with the exception of Dentist and Pilot, both of which differ from the past by no more than 2 average annual releases.
TABLE II
OFFICER RESULTS

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Air-Technical</td>
<td>22.24%</td>
<td>8.0</td>
<td>12.0</td>
<td>1.50</td>
<td>5.8</td>
<td>0.73</td>
</tr>
<tr>
<td>Cleric</td>
<td>19.35%</td>
<td>2.5</td>
<td>0.6</td>
<td>0.24</td>
<td>0.2</td>
<td>0.08</td>
</tr>
<tr>
<td>Dentist</td>
<td>12.17%</td>
<td>1.0</td>
<td>1.6</td>
<td>1.60</td>
<td>2.6</td>
<td>2.60</td>
</tr>
<tr>
<td>Elec./Comm.</td>
<td>31.53%</td>
<td>12.5</td>
<td>19.4</td>
<td>1.55</td>
<td>12.0</td>
<td>0.96</td>
</tr>
<tr>
<td>Engineering</td>
<td>29.48%</td>
<td>6.8</td>
<td>3.6</td>
<td>0.53</td>
<td>1.2</td>
<td>0.18</td>
</tr>
<tr>
<td>Flight Crew</td>
<td>28.04%</td>
<td>7.0</td>
<td>12.8</td>
<td>1.83</td>
<td>12.0</td>
<td>1.71</td>
</tr>
<tr>
<td>Lawyer</td>
<td>17.02%</td>
<td>0.3</td>
<td>0.6</td>
<td>2.40</td>
<td>0.4</td>
<td>1.60</td>
</tr>
<tr>
<td>Land-Combat Arms</td>
<td>23.88%</td>
<td>7.0</td>
<td>24.2</td>
<td>3.46</td>
<td>18.2</td>
<td>2.60</td>
</tr>
<tr>
<td>Land-Maintenance</td>
<td>31.11%</td>
<td>4.3</td>
<td>8.4</td>
<td>1.98</td>
<td>4.6</td>
<td>1.08</td>
</tr>
<tr>
<td>Logistics</td>
<td>26.73%</td>
<td>21.5</td>
<td>16.0</td>
<td>0.74</td>
<td>15.4</td>
<td>0.72</td>
</tr>
<tr>
<td>Medical</td>
<td>27.87%</td>
<td>6.3</td>
<td>10.0</td>
<td>1.60</td>
<td>7.8</td>
<td>1.25</td>
</tr>
<tr>
<td>Physician</td>
<td>25.13%</td>
<td>4.0</td>
<td>4.8</td>
<td>1.20</td>
<td>3.2</td>
<td>0.80</td>
</tr>
<tr>
<td>Pilot</td>
<td>18.40%</td>
<td>27.8</td>
<td>24.0</td>
<td>0.86</td>
<td>29.8</td>
<td>1.07</td>
</tr>
<tr>
<td>Sea-Combat/Gen.</td>
<td>19.79%</td>
<td>2.3</td>
<td>6.0</td>
<td>2.67</td>
<td>6.0</td>
<td>2.67</td>
</tr>
<tr>
<td>Security/Intelligence</td>
<td>35.90%</td>
<td>3.0</td>
<td>7.6</td>
<td>2.53</td>
<td>4.6</td>
<td>1.53</td>
</tr>
<tr>
<td>Specialist</td>
<td>27.20%</td>
<td>4.8</td>
<td>2.8</td>
<td>0.59</td>
<td>2.6</td>
<td>0.55</td>
</tr>
<tr>
<td>Sea-Technical</td>
<td>29.09%</td>
<td>3.5</td>
<td>11.4</td>
<td>3.26</td>
<td>6.2</td>
<td>1.77</td>
</tr>
<tr>
<td>All</td>
<td>29.48%</td>
<td>122.3</td>
<td>165.8</td>
<td>1.36</td>
<td>132.6</td>
<td>1.08</td>
</tr>
</tbody>
</table>

17. How significant are releases at 19 to 23 YOS compared to releases at all stages of a career? Figure 6 provides a historical depiction of attrition at 19 to 23 YOS as a fraction of total attrition. This graph shows that, for both the NCM and officer populations, attrition between 19 and 23 YOS hovered close to 10% of total attrition from 1982 to 1992 (or 1995 for officers). This fraction then increased to 35% for NCMs and 22.3% for officers in the year 2000. Attrition between 19 and 23 YOS, therefore, is expected to be a substantial component of overall attrition for the years under consideration.
IV. SUMMARY AND CONCLUSIONS

18. Manning reductions, encouraged retirement and limited recruiting in the past have led to undesirable YOS distributions, with a large proportion of CF members having between 10 and 20 YOS.

19. Recent attrition rates between 19 and 23 YOS have been higher than attrition rates during the years immediately preceding the FRP, as indeed they have been higher than attrition rates on record, dating back to 1982.

20. The large proportion of members between 10 and 20 YOS is expected to leave the CF at the more recent, higher attrition rates. During the next five years, overall NCM releases are forecast to be more than double those of the past, and officer releases are forecast to be approximately 36% greater.

21. Among all occupational groups, the Land – Combat Arms groups (both NCM and officer) appear to be in the most critical state regarding releases from 19 to 23 YOS. All NCM occupational groups are expected to improve over the second period (2007 to 2011), with the exception of Land – Combat Arms, Sea – Combat Arms/General and Sea – Technical, all of which are in the most serious category of classification for this study.

22. Attrition between 19 and 23 YOS should be considered a substantial component of overall attrition, and as such, should be given appropriate attention.
USING SATISFACTION GAPS TO EXPLORE RETENTION INTENTIONS OF MILITARY MEMBERS

John C. Helmick
Kimya S. Lee
Cynthia V. Helba
Shelley Perry
Westat, Rockville MD, 20850-3195
helmickj1@westat.com

This paper explores the relationship between satisfaction with aspects of military life and intention to leave the military. The paper builds on previous work (Hay Group, 2001) in which the satisfaction levels of employees who planned to stay in their organization were compared with those who planned to leave their organization. The Hay Group approach proposes that wide disparities in levels of satisfaction or "satisfaction gaps" between the two types of employees suggest the greatest causes of attrition.

The population of interest consists of active duty Army, Navy, Marine Corps, Air Force, and Coast Guard members who completed the 1999 Survey of Active Duty Personnel (ADS), sponsored by the Office of the Assistant Secretary of Defense for Force Management Policy and conducted by the Defense Manpower Data Center.

Using a three-step analytic approach we examine the way in which satisfaction gaps are associated with differences in military members likelihood of reporting that they intend to leave the military. In the first step, satisfaction gaps are identified. Then, we examine these gaps by Military Service, and Military Pay Grade Group. Finally, we conduct multivariate analyses to identify those gaps that have the greatest influence on retention intention. Results from these analyses provide policy makers with information that will be useful in their efforts to retain highly trained, experienced service members by presenting a focus on specific aspects of military life as possible motivators of negative career decisions.

Introduction

The U.S. military, like other organizations, constantly reviews the effects its policies and practices have on its members' career intentions as an integral part of the effort to recruit and retain quality performers. Military researchers and policy makers at all levels continuously investigate factors that influence military members' decisions to stay in or leave the military.

While there is a link between levels of satisfaction or dissatisfaction with military life and attrition and retention, it is not always clear cut. Not everyone who is satisfied intends to stay and not everyone who is dissatisfied intends to leave. Recent surveys of active duty military personnel find that while about 73 percent of those who are satisfied with the military way of life indicate they intend to stay, there are also about 20 percent of those who are dissatisfied that indicate an intention to stay.
Specific reasons for staying or leaving are unclear. When service members are asked to identify their main reasons for leaving the military and their main reasons for staying in, the opposing lists often cite identical reasons.

To perhaps better understand the linkages of military job satisfaction and career intentions, we turn to some recently publicized work by the Hay Group. In their Working Paper entitled: *The Retention Dilemma, Why Productive Workers Leave - Seven Suggestions for Keeping Them*, (2001) authors Sherman, Alper and Wolfson present a straightforward approach to examining employee attrition and its causes. Drawing on their own global survey databases of employee opinions and intentions, the authors identified two types of employee commitment groups and examined differences in levels of satisfaction with varying aspects of the job and organizational environment between these groups. The two employee groups were described as those employees stating that they planned to leave the organization within the next two years; and, those employees

---

**Top Five Reasons Cited by Enlisted Personnel for Staying In or Leaving the Military (All Enlisted Personnel)**

<table>
<thead>
<tr>
<th>Reasons for Staying</th>
<th>Percent</th>
<th>Reasons for Leaving</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Compensation</td>
<td>45</td>
<td>1 Compensation</td>
<td>51</td>
</tr>
<tr>
<td>2 Assignment Stability</td>
<td>33</td>
<td>2 Military Life</td>
<td>31</td>
</tr>
<tr>
<td>3 Military Life</td>
<td>26</td>
<td>3 Workload</td>
<td>23</td>
</tr>
<tr>
<td>4 Workload</td>
<td>15</td>
<td>4 Career Progression</td>
<td>18</td>
</tr>
<tr>
<td>5 Career Progression</td>
<td>14</td>
<td>5 Assignment Stability</td>
<td>14</td>
</tr>
</tbody>
</table>

stating that they planned to remain with the organization for more than two years. Aspects of the job and organizational environment included advancement opportunities, benefits, decision making, job autonomy, pay, recognition, supervision, teamwork, training, top management, respectful treatment, and workload. Those aspects with the largest differences in levels of satisfaction between the two commitment groups were suggested as having the greatest influence on attrition.

Survey Background

Many DOD surveys of military personnel include items that collect information on career plans and satisfaction with various aspects of military life and the military job environment. The most recent of these surveys is the 1999 Survey of Active Duty Personnel.

The 1999 Survey of Active Duty Personnel continues a line of research begun by the U.S. Department of Defense (DOD) in 1969 with a series of small-scale surveys administered approximately every two years. These surveys were expanded in 1978 to provide policymakers with information about the total population directly involved with active-duty military life. Large-scale active duty surveys were also conducted in 1985, 1992, 1993, and 1994. The 1999 survey was a 20-page, 112-question (some with multiple items) self-administering mail survey. The questionnaire included items on attitudes, experiences, career intentions, and demographics of military members.

The sample consisted of 66,040 service members drawn from a May 1999 population of 1,419,269 active duty DOD and U.S. Coast Guard personnel that were below the rank of admiral or general and had at least six months of service. Mailout data collection was conducted August 1999 through January 2000. A total of 33,189 usable surveys were returned by eligible respondents by the end of data collection resulting in a weighted response rate (corrected for non-proportional sampling) of 50.7 percent.

Identification of Satisfaction Gaps

The survey provides data particularly appropriate for an examination of satisfaction gaps as influencers of retention intentions of military members. The survey includes an item that collects data on the "retention intentions" of the respondents.

Suppose that you have to decide whether to stay on active duty. Assuming you could stay, how likely is it that you would choose to do so?

?? Very likely
?? Likely
?? Neither likely nor unlikely
?? Unlikely
?? Very unlikely
Several questions after the item regarding "retention intentions", the respondent is asked to indicate his/her level of satisfaction (i.e., Very satisfied, Satisfied, Neither satisfied nor dissatisfied, Dissatisfied, Very dissatisfied) with 37 items related to specific aspects of military life. Fourteen of the 37 items address such aspects as special and incentive pay, housing allowances, and other general military family issues. These fourteen items include a "Does not apply" response since they do not apply to all military members.

For the purpose of this analysis, we consider only two types of respondents to the item regarding the likelihood of choosing to remain on active duty:

Members choosing to stay - those members indicating that they would be "Likely" or "Very likely" to choose to stay on active duty; and,

Members choosing to leave - those members indicating that they would be "Unlikely" or "Very unlikely" to choose to stay on active duty;

**Likelihood of choosing to stay on active duty**

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Likely</td>
<td>29.1</td>
</tr>
<tr>
<td>Likely</td>
<td>22.0</td>
</tr>
<tr>
<td>Neither likely nor unlikely</td>
<td>14.1</td>
</tr>
<tr>
<td>Unlikely</td>
<td>13.8</td>
</tr>
<tr>
<td>Very Unlikely</td>
<td>20.9</td>
</tr>
</tbody>
</table>

Note: Members indicating that they would be "Neither likely nor unlikely" to choose to remain on active duty are not included in this analysis.

In addition, for each of the 37 items addressing the specific aspects of military life, we collapse the responses of "Satisfied" and "Very Satisfied" into the single measure of "percent satisfied."

**Satisfaction Gap Analysis**

To conduct the analysis we compared the percent of each respondent group (those that would chose to stay, and those that would chose to leave) that indicated they were satisfied (i.e., Satisfied or Very satisfied) with each of the items addressing the aspects of military life. Items that had the widest disparities between the satisfaction levels of the two respondent groups suggest the greatest causes of attrition.

Table 1 presents aggregated data for the two respondent groups across the 37 items addressing specific aspects of military life. For each aspect, Table 1 lists the percent of members choosing to stay who indicated satisfaction (i.e., responding
“Satisfied” or “Very satisfied”) with that aspect of military life, and the corresponding percent of members choosing to leave who indicated satisfaction with that same aspect.

Table 1 also lists the satisfaction gap for each aspect of military life. This gap is calculated by subtracting the percent of members choosing to leave from the percent of members choosing to stay for each aspect. For ease in interpreting the data, Table 1 has been sorted in descending gap order. Items high on the list (i.e., with the greatest gaps) include: Enjoyment from job; Military values, lifestyle, and tradition; Type of assignment; Quality of leadership; and, Training and professional development. Basic pay falls at almost the exact midpoint of the list, perhaps indicating not a strong

Table 1. Member Job Satisfaction and Retention Choices

<table>
<thead>
<tr>
<th>Satisfaction With: (* Items allowed a “Does not apply” response)</th>
<th>Total Percent Satisfied</th>
<th>Members Choosing to Stay</th>
<th>Members Choosing to Leave</th>
<th>Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount of enjoyment from job</td>
<td>56.6</td>
<td>28.2</td>
<td>28.4</td>
<td></td>
</tr>
<tr>
<td>Military values, lifestyle, and tradition</td>
<td>60.7</td>
<td>33.7</td>
<td>27.0</td>
<td></td>
</tr>
<tr>
<td>Type of Assignments Received</td>
<td>62.0</td>
<td>35.6</td>
<td>26.4</td>
<td></td>
</tr>
<tr>
<td>Quality of Leadership</td>
<td>46.6</td>
<td>25.8</td>
<td>20.8</td>
<td></td>
</tr>
<tr>
<td>Training and Professional Development</td>
<td>58.8</td>
<td>38.3</td>
<td>20.4</td>
<td></td>
</tr>
<tr>
<td>Frequency of PCS moves</td>
<td>44.8</td>
<td>27.1</td>
<td>17.8</td>
<td></td>
</tr>
<tr>
<td>* Co-location with military spouse</td>
<td>58.4</td>
<td>40.6</td>
<td>17.7</td>
<td></td>
</tr>
<tr>
<td>Unit morale</td>
<td>38.2</td>
<td>20.7</td>
<td>17.6</td>
<td></td>
</tr>
<tr>
<td>* Location/Station of choice, homeporting</td>
<td>53.9</td>
<td>37.2</td>
<td>16.7</td>
<td></td>
</tr>
<tr>
<td>Chances for future advancement</td>
<td>44.5</td>
<td>29.3</td>
<td>15.2</td>
<td></td>
</tr>
<tr>
<td>Amount of personal/family time</td>
<td>38.8</td>
<td>24.1</td>
<td>14.8</td>
<td></td>
</tr>
<tr>
<td>* Schools for children</td>
<td>60.0</td>
<td>45.3</td>
<td>14.7</td>
<td></td>
</tr>
<tr>
<td>Personal workload</td>
<td>46.0</td>
<td>31.9</td>
<td>14.1</td>
<td></td>
</tr>
<tr>
<td>Job security</td>
<td>77.6</td>
<td>64.0</td>
<td>13.6</td>
<td></td>
</tr>
<tr>
<td>Medical Care for Self</td>
<td>57.6</td>
<td>44.4</td>
<td>13.2</td>
<td></td>
</tr>
<tr>
<td>Deployments</td>
<td>37.0</td>
<td>23.8</td>
<td>13.2</td>
<td></td>
</tr>
<tr>
<td>Retirement Pay</td>
<td>23.9</td>
<td>11.2</td>
<td>12.7</td>
<td></td>
</tr>
<tr>
<td>Othr Mil Duties that take you away from perm duty sta</td>
<td>32.0</td>
<td>19.5</td>
<td>12.5</td>
<td></td>
</tr>
<tr>
<td>Basic Pay</td>
<td>28.4</td>
<td>16.2</td>
<td>12.2</td>
<td></td>
</tr>
<tr>
<td>Off duty educational opportunities</td>
<td>50.2</td>
<td>38.2</td>
<td>12.1</td>
<td></td>
</tr>
<tr>
<td>* Military Housing</td>
<td>31.6</td>
<td>20.1</td>
<td>11.5</td>
<td></td>
</tr>
<tr>
<td>* Youth activities on base</td>
<td>44.3</td>
<td>32.8</td>
<td>11.5</td>
<td></td>
</tr>
<tr>
<td>Dental Care for Self</td>
<td>66.6</td>
<td>55.5</td>
<td>11.2</td>
<td></td>
</tr>
<tr>
<td>* Military Family Support Programs</td>
<td>41.5</td>
<td>30.7</td>
<td>10.8</td>
<td></td>
</tr>
<tr>
<td>Availability of equip, parts and resources</td>
<td>26.9</td>
<td>17.8</td>
<td>9.1</td>
<td></td>
</tr>
<tr>
<td>* Special and Incentive Pay</td>
<td>27.5</td>
<td>18.5</td>
<td>9.0</td>
<td></td>
</tr>
<tr>
<td>* Spouse employment/career opportunities</td>
<td>36.7</td>
<td>28.4</td>
<td>8.3</td>
<td></td>
</tr>
<tr>
<td>Pace of promotions</td>
<td>36.4</td>
<td>28.6</td>
<td>7.8</td>
<td></td>
</tr>
<tr>
<td>Other Retirement Benefits</td>
<td>26.8</td>
<td>19.3</td>
<td>7.6</td>
<td></td>
</tr>
<tr>
<td>* Dental Care for family</td>
<td>38.4</td>
<td>31.0</td>
<td>7.4</td>
<td></td>
</tr>
<tr>
<td>Level of manning in unit</td>
<td>26.1</td>
<td>19.0</td>
<td>7.1</td>
<td></td>
</tr>
<tr>
<td>* Acceptable and Affordable Childcare</td>
<td>23.6</td>
<td>16.9</td>
<td>6.7</td>
<td></td>
</tr>
<tr>
<td>COLA adj to retirement pay</td>
<td>15.6</td>
<td>9.1</td>
<td>6.5</td>
<td></td>
</tr>
<tr>
<td>* Medical Care for family</td>
<td>41.5</td>
<td>35.4</td>
<td>6.1</td>
<td></td>
</tr>
<tr>
<td>* SEPRATS/COMRATS subsistence Allowance</td>
<td>29.8</td>
<td>24.5</td>
<td>5.3</td>
<td></td>
</tr>
<tr>
<td>* Re-enl Bonus/Continuation Pay</td>
<td>21.8</td>
<td>16.8</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>* Housing Allowance</td>
<td>25.6</td>
<td>20.8</td>
<td>4.8</td>
<td></td>
</tr>
</tbody>
</table>

association with retention decisions.
These results are similar to those reported by the Hay researchers who found the widest gaps in such issues as "Use of skills and abilities," "Ability of top management" and "Company has a clear sense of direction". This further supports the premise that employees are more motivated by the non-economic factors of the work environment, and that the greatest motivation is found in meaningful work derived from the desire to apply skills in a challenging effort. Employees seek to be seen as a useful and helpful member of a team guided by capable leaders who have a clear sense of direction.

Table 2 presents the satisfaction gaps for each aspect of military life for each of the five military services. To facilitate comparisons across services, Table 2 also includes the ranking of each gap (from highest to lowest) within each military service. Table 2 has been sorted by the average ranking across each of the five services.

The themes discovered in the gap analysis for all members are even clearer at the military service level. Aspects of military life with the top three largest gaps across all services include: Enjoyment from job; Military values, lifestyle, and tradition; and Type of assignment. Quality of leadership; and Unit morale round out items found in the top ten rankings across all services. Basic pay satisfaction gaps continue to remain at or below the list mid point and rank 20 and 28 for the Navy and Air Force respectively.

The greatest satisfaction gaps are observed in the Marine Corps - Enjoyment from job; Military values, lifestyle, and tradition; Type of assignment and Quality of leadership all show gaps in satisfaction between members choosing to stay and members choosing to leave of 30 percentage points or more.

Chances for future advancement shows the greatest range in satisfaction gaps across the five services; ranging from a high of 28.4 (and a rank of 7) for the Marine Corps to a low of 3.8 (and rank of 30) for the Coast Guard. Chances for future advancement rank in both the Army and Air Force top ten highest gaps, ranking at 9 and 5 respectively and showing satisfaction gaps of slightly less than 20 points in both services. Chances for future advancement for the Navy shows a satisfaction gap of only 6.5 points and a rank of 30.

Table 3 presents the satisfaction gaps for each aspect of military life for each of seven military pay grade groups. To facilitate comparisons across the pay grade groups, Table 3 also includes the ranking of each gap (from highest to lowest) within each pay grade group. Table 3 has been sorted by the average ranking across each of the seven pay grade groups. When examining data by pay grade groups it should be noted that there are no Warrant Officers (pay grade group W1-W5) in the Air Force.

Enjoyment from job, Military values, lifestyle, and tradition, Quality of leadership, Type of assignment, and Training and professional development continue to show the highest satisfaction gaps across the pay grade groups - all of these items rank in the top ten highest satisfaction gaps within each military pay grade group. Chances for future advancement for pay grade group E5-E6, with a satisfaction gap of 7.3 percentage points, did not appear to be a major consideration between members choosing to stay and members choosing to leave.
### Table 2.
Satisfaction Gaps in Aspects of Military Life by Service

<table>
<thead>
<tr>
<th>Aspects of Military Life</th>
<th>Army Gap Rank</th>
<th>Army Gap Rank</th>
<th>Air Force Gap Rank</th>
<th>Air Force Gap Rank</th>
<th>Marine Corps Gap Rank</th>
<th>Marine Corps Gap Rank</th>
<th>Coast Guard Gap Rank</th>
<th>Coast Guard Gap Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount of enjoyment from job</td>
<td>28.2 1</td>
<td>27.3 1</td>
<td>24.6 2</td>
<td>36.4 2</td>
<td>32.3 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Military values, lifestyle, and tradition</td>
<td>25.8 3</td>
<td>25.4 2</td>
<td>27.2 1</td>
<td>36.8 1</td>
<td>26.0 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of Assignments Received</td>
<td>26.2 2</td>
<td>25.3 3</td>
<td>24.5 3</td>
<td>32.1 3</td>
<td>21.1 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality of Leadership</td>
<td>19.1 6</td>
<td>21.7 4</td>
<td>19.8 4</td>
<td>30.1 4</td>
<td>13.4 9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit morale</td>
<td>19.3 5</td>
<td>15.6 9</td>
<td>15.8 9</td>
<td>24.1 8</td>
<td>14.1 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training and Professional Development</td>
<td>22.3 4</td>
<td>18.1 6</td>
<td>17.7 8</td>
<td>28.0 6</td>
<td>10.5 14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Co-location with military spouse</td>
<td>15.4 11</td>
<td>18.4 5</td>
<td>10.9 19</td>
<td>28.4 5</td>
<td>27.8 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency of PCS moves</td>
<td>15.7 10</td>
<td>15.7 8</td>
<td>18.6 6</td>
<td>23.3 9</td>
<td>12.6 12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Location/Station of choice, homeporting</td>
<td>17.6 8</td>
<td>14.4 11</td>
<td>18.0 7</td>
<td>19.5 11</td>
<td>6.8 23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Schools for children</td>
<td>18.2 7</td>
<td>12.1 16</td>
<td>11.1 17</td>
<td>17.7 14</td>
<td>11.5 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal workload</td>
<td>14.1 15</td>
<td>12.1 15</td>
<td>15.2 10</td>
<td>18.6 13</td>
<td>10.4 15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job security</td>
<td>13.6 17</td>
<td>11.1 19</td>
<td>13.2 13</td>
<td>20.0 10</td>
<td>13.1 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amount of personal/family time</td>
<td>14.7 14</td>
<td>12.5 13</td>
<td>13.5 11</td>
<td>19.3 12</td>
<td>6.9 22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deployments</td>
<td>13.5 18</td>
<td>13.6 12</td>
<td>12.8 14</td>
<td>16.0 17</td>
<td>8.9 18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chances for future advancement</td>
<td>17.1 9</td>
<td>6.5 30</td>
<td>19.3 5</td>
<td>25.9 7</td>
<td>3.8 30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical Care for Self</td>
<td>11.6 22</td>
<td>16.3 7</td>
<td>11.6 16</td>
<td>17.3 15</td>
<td>6.3 24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Othr Mil Duties that take you away from perm duty sta</td>
<td>13.9 16</td>
<td>11.7 17</td>
<td>11.0 18</td>
<td>15.9 18</td>
<td>7.6 20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retirement Pay</td>
<td>12.6 19</td>
<td>11.5 18</td>
<td>13.2 12</td>
<td>15.1 21</td>
<td>6.9 21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Youth activities on base</td>
<td>15.0 12</td>
<td>7.6 29</td>
<td>8.6 23</td>
<td>15.1 22</td>
<td>13.6 8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Military Family Support Programs</td>
<td>10.0 25</td>
<td>12.1 14</td>
<td>8.0 26</td>
<td>13.6 25</td>
<td>14.6 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic Pay</td>
<td>14.9 13</td>
<td>10.8 20</td>
<td>7.8 28</td>
<td>15.6 19</td>
<td>9.3 17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dental Care for Self</td>
<td>9.7 26</td>
<td>15.3 10</td>
<td>6.6 30</td>
<td>15.1 20</td>
<td>9.7 16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Military Housing</td>
<td>11.9 21</td>
<td>9.0 24</td>
<td>12.0 15</td>
<td>12.9 26</td>
<td>7.7 19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Off duty educational opportunities</td>
<td>12.2 20</td>
<td>10.4 21</td>
<td>9.8 22</td>
<td>16.8 16</td>
<td>1.8 33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Dental Care for family</td>
<td>5.4 35</td>
<td>9.5 23</td>
<td>6.8 29</td>
<td>11.9 28</td>
<td>14.0 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Special and Incentive Pay</td>
<td>10.4 23</td>
<td>7.9 27</td>
<td>4.6 33</td>
<td>14.6 24</td>
<td>5.5 25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pace of promotions</td>
<td>7.7 29</td>
<td>5.6 32</td>
<td>10.1 20</td>
<td>14.8 23</td>
<td>3.6 31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Availability of equip, parts and resources</td>
<td>10.0 24</td>
<td>7.7 28</td>
<td>8.4 24</td>
<td>8.2 33</td>
<td>4.7 27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Retirement Benefits</td>
<td>6.0 34</td>
<td>9.5 22</td>
<td>7.9 27</td>
<td>12.6 27</td>
<td>5.0 26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Medical Care for family</td>
<td>6.2 33</td>
<td>6.2 31</td>
<td>6.0 31</td>
<td>6.4 35</td>
<td>12.9 11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Spouse employment/career opportunities</td>
<td>8.2 28</td>
<td>8.5 25</td>
<td>8.0 25</td>
<td>8.2 32</td>
<td>0.2 36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Acceptable and Affordable Childcare</td>
<td>7.4 30</td>
<td>8.0 26</td>
<td>2.8 36</td>
<td>10.0 30</td>
<td>3.8 29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level of manning in unit</td>
<td>8.3 27</td>
<td>4.5 36</td>
<td>9.8 21</td>
<td>6.0 36</td>
<td>3.3 32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COLA adj to retirement pay</td>
<td>6.7 31</td>
<td>5.5 33</td>
<td>5.9 32</td>
<td>10.5 28</td>
<td>1.8 34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Re-enl Bonus/Continuation Pay</td>
<td>5.0 36</td>
<td>5.0 34</td>
<td>2.6 37</td>
<td>7.8 34</td>
<td>4.6 28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* SEPRATS/COMRATS subsistence Allowance</td>
<td>5.0 37</td>
<td>4.8 35</td>
<td>4.1 34</td>
<td>9.9 31</td>
<td>1.6 35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Housing Allowance</td>
<td>6.5 32</td>
<td>3.7 37</td>
<td>4.0 35</td>
<td>5.5 37</td>
<td>-0.5 37</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Multivariate Analysis

Multivariate analysis is the next step after examining the satisfaction gap for each individual item. The satisfaction gap analysis allows one to examine the relative association of each of the 37 aspects of military life with the reported retention intention of military personnel. The advantage of using multivariate analyses when studying retention intention is that the researcher can quantify unique effects of the independent variables. For this demonstration logistic regression analysis was utilized.

Logistic regression is a variant of linear regression; it assesses the relationship between the criterion variable (i.e., members who reported being likely to stay in the military vs. members reporting being unlikely to stay) and several predictor variables (i.e., members' satisfaction with types of assignments received). Logistic regression allows the researcher to estimate the odds of an event occurring (i.e., the odds of staying in the military) on the basis of the values for the predictor variables.

The following steps were utilized in predicting members' likelihood of staying in the military. We took the top 1/3 of items with the largest satisfaction gap between those
who reported being likely/very likely to stay in the military and those who reported being unlikely/very unlikely to stay in the military. The top items were:

- Amount of enjoyment from job
- Military values, lifestyle, and tradition
- Type of Assignments Received
- Quality of Leadership
- Training and Professional Development
- Frequency of PCS moves
- Co-location with military spouse
- Unit morale
- Location/Station of choice homeporting
- Chances for future advancement
- Amount of personal/family time
- Schools for children

In addition to these 12 items, we also included Service, paygrade, and whether the member had any children or legal dependents in the model. All variables in the model were recoded to a dichotomous variable (i.e., amount of enjoyment from job -- 1 = satisfied; 0 = all other responses). A forward logistic regression was conducted to determine which of the 12 independent variables were predictors of military members’ retention intentions. An initial analysis was first conducted on all variables. This model was used to give us a better understand which variables are most significant in predicting intention to stay in the military. Based on this initial model, the variables, training and professional development, co-location with military spouse, unit morale, and location/station of choice, homeporting were excluded from the final model because they did not have a significant effect on the criterion variable.

We controlled for Service, paygrade, and children in predicting the odds of intention to stay in the military. Preliminary analyses indicated that there is some variation by Service and paygrade for the aspects of military life with the largest gaps (see Table 2 and Table 3). These key demographic characteristics therefore had to be accounted for and were included in the analysis.

Table 4 shows the results of the logistic regression analysis. The results indicated that the odds of intending to stay in the military for members who are satisfied with the amount of enjoyment from their jobs are 65 percent higher than the odds for members who are not satisfied, controlling for other variables in the model. The odds of intending to stay in the military for members who are satisfied with the military values, lifestyle, and traditions are 59 percent higher than the odds for members who are not satisfied.
Table 4.
Summary of logistic regression analysis predicting staying in the military

<table>
<thead>
<tr>
<th>Predictor</th>
<th>β</th>
<th>SE</th>
<th>Odds ratio</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.63</td>
<td>.084</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Satisfaction Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amount of enjoyment from job</td>
<td>.50**</td>
<td>.041</td>
<td>1.65</td>
<td>1.52</td>
<td>1.79</td>
</tr>
<tr>
<td>Military values, lifestyle, and tradition</td>
<td>.47**</td>
<td>.040</td>
<td>1.59</td>
<td>1.47</td>
<td>1.73</td>
</tr>
<tr>
<td>Chances for future advancement</td>
<td>.32**</td>
<td>.038</td>
<td>1.37</td>
<td>1.27</td>
<td>1.48</td>
</tr>
<tr>
<td>Type of Assignments Received</td>
<td>.28**</td>
<td>.041</td>
<td>1.32</td>
<td>1.22</td>
<td>1.44</td>
</tr>
<tr>
<td>Quality of Leadership</td>
<td>.25**</td>
<td>.048</td>
<td>1.28</td>
<td>1.16</td>
<td>1.41</td>
</tr>
<tr>
<td>Amount of personal/family time</td>
<td>.21**</td>
<td>.042</td>
<td>1.24</td>
<td>1.14</td>
<td>1.34</td>
</tr>
<tr>
<td>Schools for children</td>
<td>.14**</td>
<td>.049</td>
<td>1.15</td>
<td>1.05</td>
<td>1.27</td>
</tr>
<tr>
<td>Frequency of PCS moves</td>
<td>.11*</td>
<td>.039</td>
<td>1.11</td>
<td>1.03</td>
<td>1.20</td>
</tr>
<tr>
<td>Demographic variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E1-E3</td>
<td>-1.25**</td>
<td>.069</td>
<td>0.29</td>
<td>0.25</td>
<td>0.33</td>
</tr>
<tr>
<td>E4</td>
<td>-.95**</td>
<td>.063</td>
<td>0.39</td>
<td>0.34</td>
<td>0.44</td>
</tr>
<tr>
<td>E5-E6</td>
<td>.03</td>
<td>.042</td>
<td>1.03</td>
<td>0.95</td>
<td>1.12</td>
</tr>
<tr>
<td>E7-E9</td>
<td>-.24**</td>
<td>.047</td>
<td>0.78</td>
<td>0.71</td>
<td>0.86</td>
</tr>
<tr>
<td>W1-W5</td>
<td>.05</td>
<td>.074</td>
<td>1.05</td>
<td>0.91</td>
<td>1.22</td>
</tr>
<tr>
<td>O1-O3</td>
<td>-.63**</td>
<td>.045</td>
<td>0.53</td>
<td>0.49</td>
<td>0.58</td>
</tr>
<tr>
<td>Army</td>
<td>-.02</td>
<td>.075</td>
<td>0.98</td>
<td>0.84</td>
<td>1.14</td>
</tr>
<tr>
<td>Navy</td>
<td>.03</td>
<td>.071</td>
<td>1.04</td>
<td>0.90</td>
<td>1.19</td>
</tr>
<tr>
<td>Marine Corps</td>
<td>-.22**</td>
<td>.075</td>
<td>0.81</td>
<td>0.70</td>
<td>0.93</td>
</tr>
<tr>
<td>Air Force</td>
<td>.15*</td>
<td>.071</td>
<td>1.17</td>
<td>1.01</td>
<td>1.34</td>
</tr>
<tr>
<td>Children</td>
<td>.42**</td>
<td>.038</td>
<td>1.52</td>
<td>1.41</td>
<td>1.64</td>
</tr>
</tbody>
</table>

*p < .01; **p < .001

Interpreting coefficients in terms of odds ratios is not always straightforward -- odds ratios may not always be meaningful to leaders and policymakers. Indicating that the odds of staying in the military for members who report being satisfied with certain aspects of military life does not indicate whether this satisfaction is meaningful. It does not indicate what the base point is. Interpreting the logistic model in terms of estimated probability might be more useful.

From the model above, we can calculate the probability that a member will report that he or she intends to stay in the military. For example, suppose the profile of our military member was Air Force, E7-E9, with children, and not satisfied with any of the 8 aspects of military life included in our model. Based on this information, we would discover that this member has a 24 percent probability of reporting that he or she intends to stay in the military compared with a 76 percent probability when he or she is satisfied with all 8 aspects of military life. If a member is satisfied with all 8 aspects of military...
life included in the model, their estimated probability of intending to stay in the military increases by approximately 52 percent.\textsuperscript{30}

A less dramatic example is illustrated with Marine Corps, E1-E3, without children, and he or she was not satisfied with any of the 8 aspects of military life included in the model. Based this profile the member a 11 percent probability of reporting that he or she intends to stay in the military compared with a 44 percent probability when he or she is satisfied with all 8 aspects of military life.

What about the individual items? Taking that same Air Force, E7-E9, with children, if he or she is satisfied with the amount of enjoyment from his or her job, this one item will increase the estimated probability of reporting that her or he intends to stay in the military by 10 percent to a 35 percent estimated probability that he or she will stay the military.

**Conclusion**

This paper is intended to provide a general approach that DOD policymakers and those in the individual Services can utilize when studying the retention intentions of military members. The results of the analysis adds information on the magnitude of the effects attitudinal variables have on stated retention intention. We have provided a tool for the individual Services to model the effect of different types of satisfaction, such as leadership quality or frequency of PCS moves, on different groups of military members. This information may be used in the effort to more clearly understand what retains officers and enlisted members in the military.

**REFERENCE**


\textsuperscript{30} For a discussion of logistic models in terms of probabilities see Long, 1997.
CFAT / AAT CONCORDANCE STUDY: A PRELIMINARY INVESTIGATION OF THE UTILITY OF THE AAT AS A PROXY FOR THE CFAT

John J. Donohue, Ph.D.
Director Human Resources Research & Evaluation
National Defence Headquarters, Ottawa, Canada K1A 0K2
donohue.jj@forces.ca


EXECUTIVE SUMMARY

The purpose of this report is to determine the conversion table for equating raw scores on the Australian Aptitude Test (AAT) to raw scores on the Canadian Forces Aptitude Test (CFAT). The psychometric analyses of the two tests and their respective subscales consisted of explorations of the scale reliabilities, item difficulties, gender bias, distractors, and the construction of the concordance table. The scale reliabilities were fairly good, with only the VS subscale of the CFAT raising concerns. The item difficulty analyses revealed some discrepancies with the expected outcomes. The gender bias (DIF) analyses revealed that gender bias appeared to be an issue on only one question on the WA subscale of the AAT. Finally, the conversion table for converting raw scores on the AAT to raw scores on the CFAT was calculated.
TABLE OF CONTENTS

EXECUTIVE SUMMARY ................................................................. 1

TABLE OF CONTENTS ................................................................. 2

LIST OF TABLES .............................................................................. 3

DATASET

DEMOGRAPHICS ........................................................................... 4

PSYCHOMETRIC ANALYSES .......................................................... 5

PSYCHOMETRIC FINDINGS ............................................................ 5

Order Effect Analysis .................................................................. 5

Reliability Analyses ................................................................... 7

Item Difficulties .......................................................................... 8

Item Gender Bias Analyses ......................................................... 14

Distractor Analyses .................................................................. 14

Test Concordance ...................................................................... 20

CONCLUSIONS .............................................................................. 25

REFERENCES .................................................................................. 27
# List of Tables

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Table 1. Means, Standard Deviations and Sample Sizes for the CFAT and AAT Tests and Subscales</td>
</tr>
<tr>
<td>7</td>
<td>Table 2. Means, Standard Deviations and Sample Sizes for the CFAT and AAT Subscales by Administration Date</td>
</tr>
<tr>
<td>9</td>
<td>Table 3. Alpha Reliabilities and Sample Sizes for the CFAT and AT Tests and Subscales</td>
</tr>
<tr>
<td>10</td>
<td>Table 4. Item Difficulties for the CFAT Subscales</td>
</tr>
<tr>
<td>12</td>
<td>Table 5. Item Difficulties for the AAT Subscales</td>
</tr>
<tr>
<td>15</td>
<td>Table 6. P-values for the 2df Chi-squared DIF Analysis for Females versus Males on the CFAT Items</td>
</tr>
<tr>
<td>17</td>
<td>Table 7. P-values for the 2df Chi-squared DIF Analysis for Females versus Males on the AAT Items</td>
</tr>
<tr>
<td>19</td>
<td>Table 8. WLS R-squared Calculations for Uniform and Non-uniform DIF (Zumbo, 1999)</td>
</tr>
<tr>
<td>21</td>
<td>Table 9. Item Responses for CFAT Items</td>
</tr>
<tr>
<td>23</td>
<td>Table 10. Item Responses for AAT Items</td>
</tr>
<tr>
<td>26</td>
<td>Table 11. Table for Converting AAT Total Score to CFAT Total Score</td>
</tr>
</tbody>
</table>
**CFAT / AAT Concordance Study: A Preliminary Investigation**

*of the Utility of the AAT as a Proxy for the CFAT*

**DATASET DEMOGRAPHICS**

The purpose of this study was to determine the conversion table, or concordance table, for converting raw total scores on the Australian Aptitude Test (AAT) to raw total scores on the Canadian Forces Aptitude Test (CFAT). This table allows the Canadian Forces to use the AAT as an interim assessment of candidate’s ability should the CFAT be compromised in some way.

The CFAT is a timed test comprised of three subscales designed to assess Verbal Skill (15 items), Spatial Awareness (15 items) and Problem Solving ability (30 items). The AAT is also a timed test comprised of three subscales. The MX subscale (28 items) is designed to assess Mathematical Reasoning (analogous to the PS subscale of the CFAT), while the WA subscale (30 items) can be compared to the Verbal Skills subscale. The third AAT subscale addresses Clerical abilities, and thus has no analogous CFAT subscale. The data collected for the Clerical subscale of the AAT was not used in this study. The participants in the study completed the subscale as per the administration instructions of the AAT, but the data was neither transcribed nor analyzed.

Data was collected from participants in four sessions held at a local university campus. The participants ranged in age from 18 to 29 (mean = 21.02, SD = 2.07). Although approximately equal numbers of males and females were recruited, there was a much larger turnout of female participants than males. Of the 190 participants, 124 were female and 66 were male.

Univariate descriptive analyses revealed that the male participants scored significantly higher than the female participants on the CFAT and AAT raw scales. To alleviate the impact of this discrepancy, the male and female participants were matched.

---

31 The data collected for the Clerical subscale of the AAT was not used in this study. The participants in the study completed the subscale as per the administration instructions of the AAT, but the data was neither transcribed nor analyzed.
on three criteria (CFAT raw score, AAT raw score, and age, in that order) in order to provide comparable data to explore the variables of interest in this study. This matching process yielded a data set of 66 (sixty-six) male-female pairings. The age range for this sample was from 18 to 29 (mean = 21.09, SD = 2.10). The mean scores for the participants (N =132) on each test and subscale are presented in Table 1. Visual inspection of these scores indicates that, in general, the participants in the present study scored marginally higher than did participants in the Zumbo and Hubley (1998) study on each of the three CFAT subscales. Comparison data for the AAT were not available.

PSYCHOMETRIC ANALYSES

The following psychometric analyses and tasks were performed:

1. Order Effect Analysis
2. Reliability Analyses of the CFAT and AAT
3. Item Difficulty Analysis
4. Item Gender Bias Analyses
5. Distractor Analyses
6. Test Concordance

PSYCHOMETRIC FINDINGS

Order Effect Analysis

The raw test and subscale scores for each day the testing was done (Monday, Tuesday, Wednesday and Thursday) were compared to determine whether the order in which the tests were presented made a difference in the observed responses. Participants on Monday and Thursday evenings received the AAT first and the CFAT second, while participants on Tuesday and Wednesday evenings received the CFAT first and the AAT second (these orderings were determined by random coin flip). The means and standard deviations for this analysis are presented in Table 2.
Table 1
Means, Standard Deviations and Sample Sizes for the CFAT and AAT Tests and Subscales

<table>
<thead>
<tr>
<th>Test</th>
<th>Subscale</th>
<th>Gender</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFAT</td>
<td>Total (60 items)</td>
<td>Total</td>
<td>43.83</td>
<td>6.85</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>43.65</td>
<td>6.77</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>44.00</td>
<td>6.97</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Verbal Skills (15 items)</td>
<td>Total</td>
<td>10.27</td>
<td>2.52</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>10.44</td>
<td>2.41</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>10.11</td>
<td>2.64</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Spatial Awareness (15 items)</td>
<td>Total</td>
<td>10.29</td>
<td>2.75</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>10.41</td>
<td>2.40</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>10.17</td>
<td>3.07</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Problem Solving (30 items)</td>
<td>Total</td>
<td>23.27</td>
<td>4.77</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>22.80</td>
<td>4.97</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>23.73</td>
<td>4.54</td>
<td>66</td>
</tr>
<tr>
<td>AAT</td>
<td>Total (58 items)</td>
<td>Total</td>
<td>41.06</td>
<td>6.78</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>41.21</td>
<td>6.37</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>40.91</td>
<td>7.21</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>MX (28 items)</td>
<td>Total</td>
<td>19.98</td>
<td>4.36</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>19.82</td>
<td>4.11</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>20.14</td>
<td>4.63</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>WA (30 items)</td>
<td>Total</td>
<td>21.08</td>
<td>4.01</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>21.39</td>
<td>3.91</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>20.77</td>
<td>4.10</td>
<td>66</td>
</tr>
</tbody>
</table>

One-way ANOVA tests using the subscale scores as the independent variables and administration date as the fixed factor revealed no significant differences between the
participants’ scores across the administration dates on any of the subscale scores. This finding provides evidence that there was no order effect for the administrations.

Table 2

Means, Standard Deviations and Sample Sizes for the CFAT and AAT Subscales by Administration Date

<table>
<thead>
<tr>
<th></th>
<th>Monday (n = 39)</th>
<th>Tuesday (n = 34)</th>
<th>Wednesday (n = 34)</th>
<th>Thursday (n = 25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFAT – VS subscale</td>
<td>10.31 (2.64)</td>
<td>9.85 (2.23)</td>
<td>10.50 (2.84)</td>
<td>10.48 (2.33)</td>
</tr>
<tr>
<td>CFAT - SA subscale</td>
<td>10.08 (3.16)</td>
<td>10.03 (2.71)</td>
<td>10.15 (2.11)</td>
<td>11.16 (2.87)</td>
</tr>
<tr>
<td>CFAT - PS subscale</td>
<td>22.33 (5.08)</td>
<td>22.60 (4.49)</td>
<td>24.32 (4.80)</td>
<td>24.24 (4.37)</td>
</tr>
<tr>
<td>AAT - MX subscale</td>
<td>18.49 (4.90)</td>
<td>20.59 (4.03)</td>
<td>21.15 (4.19)</td>
<td>19.88 (3.63)</td>
</tr>
<tr>
<td>AAT - WA subscale</td>
<td>20.13 (4.19)</td>
<td>21.00 (3.81)</td>
<td>22.15 (3.65)</td>
<td>21.24 (4.27)</td>
</tr>
</tbody>
</table>

Reliability Analyses

The CFAT subscales and the AAT subscales were analyzed separately, and the results are reported for both the overall data set (N =132) and the male and female participant groups (n = 66 for each group). The observed alpha reliability coefficients for the CFAT and the AAT tests and subscales are presented in Table 3.

While most of the scales are fairly consistent (α > 0.7), there were some areas that bear further examination. Specifically, the Verbal Skill subscale of the CFAT, where the
a-coefficients were quite low (in the 0.5 range). Further examination of this subscale may reveal the cause of this low observed reliability.\footnote{It should be noted that, due to the confidential nature of the CFAT and AAT tests, the researcher did not have access to validity and reliability data for the CFAT and the AAT.}

\textit{Item Difficulties}

The item difficulties for the CFAT and the AAT are presented in Tables 4 and 5, respectively. The item difficulties are presented for each gender as well as the entire matched sample.

As shown in Table 4, the item difficulties for the Spatial Awareness and Problem Solving subscales seem to approximate a monotonically decreasing linear function, with a few exceptions. This pattern is as expected in a test of this sort where the more difficult items are presented later. In addition, the fact that the tests are carefully timed would also tend to produce responses in the pattern observed.

The pattern for the Verbal Skills items, however, does not seem to match the pattern of the other two subscales. The difficulty for the first item is high, and there does not seem to be the same linear pattern among the other items. This lack of a pattern raises some concern about the observed response pattern.

No information about the AAT was available in order to make informed judgment about the test or the subscales. Further examination of the expected and observed patterns may reveal concerns with the AAT or its administration.
### Table 3

**Alpha Reliabilities and Sample Sizes for the CFAT and AAT Tests and Subscales**

<table>
<thead>
<tr>
<th>Test</th>
<th>Subscale</th>
<th>Gender</th>
<th>Reliability (a)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFAT</td>
<td>Total (60 items)</td>
<td>Total</td>
<td>.7621</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>.7684</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>.7485</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>VS (15 items)</td>
<td>Total</td>
<td>.5701</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>.5568</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>.5877</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>SA (15 items)</td>
<td>Total</td>
<td>.7225</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>.6538</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>.7676</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>PS (30 items)</td>
<td>Total</td>
<td>.7923</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>.8371</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>.7059</td>
<td>49</td>
</tr>
<tr>
<td>AAT</td>
<td>Total (58 items)</td>
<td>Total</td>
<td>.8171</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>.7942</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>.8392</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>MX (28 items)</td>
<td>Total</td>
<td>.7869</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>.7602</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>.8124</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>WA (30 items)</td>
<td>Total</td>
<td>.7462</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>.7367</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>.7589</td>
<td>66</td>
</tr>
</tbody>
</table>
Table 4

*Item Difficulties for the CFAT Subscales*

<table>
<thead>
<tr>
<th>CFAT Subscale</th>
<th>Item</th>
<th>Total</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Skill</td>
<td>VS1</td>
<td>.5862</td>
<td>.5000</td>
<td>.6596</td>
</tr>
<tr>
<td></td>
<td>VS2</td>
<td>.8736</td>
<td>.9250</td>
<td>.8298</td>
</tr>
<tr>
<td></td>
<td>VS3</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>VS4</td>
<td>.7816</td>
<td>.8000</td>
<td>.7660</td>
</tr>
<tr>
<td></td>
<td>VS5</td>
<td>.5632</td>
<td>.5250</td>
<td>.5957</td>
</tr>
<tr>
<td></td>
<td>VS6</td>
<td>.4828</td>
<td>.4500</td>
<td>.5106</td>
</tr>
<tr>
<td></td>
<td>VS7</td>
<td>.6782</td>
<td>.7750</td>
<td>.5957</td>
</tr>
<tr>
<td></td>
<td>VS8</td>
<td>.8161</td>
<td>.8500</td>
<td>.7872</td>
</tr>
<tr>
<td></td>
<td>VS9</td>
<td>.8161</td>
<td>.8250</td>
<td>.8085</td>
</tr>
<tr>
<td></td>
<td>VS10</td>
<td>.7701</td>
<td>.8500</td>
<td>.7021</td>
</tr>
<tr>
<td></td>
<td>VS11</td>
<td>.7586</td>
<td>.7250</td>
<td>.7872</td>
</tr>
<tr>
<td></td>
<td>VS12</td>
<td>.5172</td>
<td>.5250</td>
<td>.5106</td>
</tr>
<tr>
<td></td>
<td>VS13</td>
<td>.5402</td>
<td>.5000</td>
<td>.5745</td>
</tr>
<tr>
<td></td>
<td>VS14</td>
<td>.5747</td>
<td>.5250</td>
<td>.6170</td>
</tr>
<tr>
<td></td>
<td>VS15</td>
<td>.4368</td>
<td>.4000</td>
<td>.4681</td>
</tr>
<tr>
<td>Spatial Awareness</td>
<td>SA1</td>
<td>.8851</td>
<td>.9000</td>
<td>.8723</td>
</tr>
<tr>
<td></td>
<td>SA2</td>
<td>.9425</td>
<td>.9250</td>
<td>.9574</td>
</tr>
<tr>
<td></td>
<td>SA3</td>
<td>.7816</td>
<td>.7500</td>
<td>.8085</td>
</tr>
<tr>
<td></td>
<td>SA4</td>
<td>.9310</td>
<td>.9250</td>
<td>.9362</td>
</tr>
<tr>
<td></td>
<td>SA5</td>
<td>.9080</td>
<td>.9000</td>
<td>.9149</td>
</tr>
<tr>
<td></td>
<td>SA6</td>
<td>.7931</td>
<td>.8250</td>
<td>.7660</td>
</tr>
<tr>
<td></td>
<td>SA7</td>
<td>.7816</td>
<td>.8000</td>
<td>.7660</td>
</tr>
<tr>
<td></td>
<td>SA8</td>
<td>.7816</td>
<td>.7500</td>
<td>.8085</td>
</tr>
<tr>
<td></td>
<td>SA9</td>
<td>.7816</td>
<td>.8000</td>
<td>.7660</td>
</tr>
<tr>
<td></td>
<td>SA10</td>
<td>.6552</td>
<td>.7250</td>
<td>.5957</td>
</tr>
<tr>
<td></td>
<td>SA11</td>
<td>.5977</td>
<td>.5250</td>
<td>.6596</td>
</tr>
<tr>
<td></td>
<td>SA12</td>
<td>.4253</td>
<td>.3750</td>
<td>.4681</td>
</tr>
<tr>
<td></td>
<td>SA13</td>
<td>.3678</td>
<td>.2750</td>
<td>.4468</td>
</tr>
<tr>
<td></td>
<td>SA14</td>
<td>.3793</td>
<td>.3750</td>
<td>.3830</td>
</tr>
<tr>
<td></td>
<td>SA15</td>
<td>.3678</td>
<td>.2750</td>
<td>.4468</td>
</tr>
</tbody>
</table>
Table 4 continued

<table>
<thead>
<tr>
<th>CFAT Subscale</th>
<th>Item</th>
<th>Total</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem Solving</td>
<td>PS1</td>
<td>.9310</td>
<td>.9250</td>
<td>.9362</td>
</tr>
<tr>
<td></td>
<td>PS2</td>
<td>.9425</td>
<td>.9250</td>
<td>.9574</td>
</tr>
<tr>
<td></td>
<td>PS3</td>
<td>.9080</td>
<td>.8750</td>
<td>.9362</td>
</tr>
<tr>
<td></td>
<td>PS4</td>
<td>.9080</td>
<td>.9000</td>
<td>.9149</td>
</tr>
<tr>
<td></td>
<td>PS5</td>
<td>.9540</td>
<td>.9500</td>
<td>.9574</td>
</tr>
<tr>
<td></td>
<td>PS6</td>
<td>.8276</td>
<td>.8000</td>
<td>.8511</td>
</tr>
<tr>
<td></td>
<td>PS7</td>
<td>.8391</td>
<td>.7250</td>
<td>.9362</td>
</tr>
<tr>
<td></td>
<td>PS8</td>
<td>.9655</td>
<td>.9500</td>
<td>.9787</td>
</tr>
<tr>
<td></td>
<td>PS9</td>
<td>.8161</td>
<td>.8000</td>
<td>.8298</td>
</tr>
<tr>
<td></td>
<td>PS10</td>
<td>.9080</td>
<td>.8750</td>
<td>.9362</td>
</tr>
<tr>
<td></td>
<td>PS11</td>
<td>.8966</td>
<td>.9000</td>
<td>.8936</td>
</tr>
<tr>
<td></td>
<td>PS12</td>
<td>.8276</td>
<td>.8250</td>
<td>.8298</td>
</tr>
<tr>
<td></td>
<td>PS13</td>
<td>.8966</td>
<td>.8750</td>
<td>.9149</td>
</tr>
<tr>
<td></td>
<td>PS14</td>
<td>.8276</td>
<td>.8250</td>
<td>.8298</td>
</tr>
<tr>
<td></td>
<td>PS15</td>
<td>.7471</td>
<td>.7000</td>
<td>.7872</td>
</tr>
<tr>
<td></td>
<td>PS16</td>
<td>.7471</td>
<td>.7500</td>
<td>.7447</td>
</tr>
<tr>
<td></td>
<td>PS17</td>
<td>.8506</td>
<td>.8500</td>
<td>.8511</td>
</tr>
<tr>
<td></td>
<td>PS18</td>
<td>.8276</td>
<td>.8000</td>
<td>.8511</td>
</tr>
<tr>
<td></td>
<td>PS19</td>
<td>.7011</td>
<td>.6250</td>
<td>.7660</td>
</tr>
<tr>
<td></td>
<td>PS20</td>
<td>.8966</td>
<td>.8500</td>
<td>.9362</td>
</tr>
<tr>
<td></td>
<td>PS21</td>
<td>.8621</td>
<td>.8750</td>
<td>.8511</td>
</tr>
<tr>
<td></td>
<td>PS22</td>
<td>.7011</td>
<td>.6250</td>
<td>.7660</td>
</tr>
<tr>
<td></td>
<td>PS23</td>
<td>.8046</td>
<td>.7750</td>
<td>.8298</td>
</tr>
<tr>
<td></td>
<td>PS24</td>
<td>.7931</td>
<td>.7500</td>
<td>.8298</td>
</tr>
<tr>
<td></td>
<td>PS25</td>
<td>.8161</td>
<td>.7750</td>
<td>.8511</td>
</tr>
<tr>
<td></td>
<td>PS26</td>
<td>.7701</td>
<td>.6750</td>
<td>.8511</td>
</tr>
<tr>
<td></td>
<td>PS27</td>
<td>.6092</td>
<td>.7000</td>
<td>.5319</td>
</tr>
<tr>
<td></td>
<td>PS28</td>
<td>.5747</td>
<td>.5000</td>
<td>.6383</td>
</tr>
<tr>
<td></td>
<td>PS29</td>
<td>.6092</td>
<td>.6000</td>
<td>.6170</td>
</tr>
<tr>
<td></td>
<td>PS30</td>
<td>.6897</td>
<td>.6250</td>
<td>.7447</td>
</tr>
</tbody>
</table>
Table 5

<table>
<thead>
<tr>
<th>AAT Subscale</th>
<th>Item</th>
<th>Total</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>MX</td>
<td>MX1</td>
<td>.9394</td>
<td>.9697</td>
<td>.9091</td>
</tr>
<tr>
<td></td>
<td>MX2</td>
<td>.8258</td>
<td>.8485</td>
<td>.8030</td>
</tr>
<tr>
<td></td>
<td>MX3</td>
<td>.8485</td>
<td>.8333</td>
<td>.8636</td>
</tr>
<tr>
<td></td>
<td>MX4</td>
<td>.9394</td>
<td>.9545</td>
<td>.9242</td>
</tr>
<tr>
<td></td>
<td>MX5</td>
<td>.5303</td>
<td>.4242</td>
<td>.6364</td>
</tr>
<tr>
<td></td>
<td>MX6</td>
<td>.8939</td>
<td>.9091</td>
<td>.8788</td>
</tr>
<tr>
<td></td>
<td>MX7</td>
<td>.9318</td>
<td>.9394</td>
<td>.9242</td>
</tr>
<tr>
<td></td>
<td>MX8</td>
<td>.7576</td>
<td>.7424</td>
<td>.7727</td>
</tr>
<tr>
<td></td>
<td>MX9</td>
<td>.7803</td>
<td>.7576</td>
<td>.8030</td>
</tr>
<tr>
<td></td>
<td>MX10</td>
<td>.9470</td>
<td>.9394</td>
<td>.9545</td>
</tr>
<tr>
<td></td>
<td>MX11</td>
<td>.9848</td>
<td>1.0000</td>
<td>.9697</td>
</tr>
<tr>
<td></td>
<td>MX12</td>
<td>.9242</td>
<td>.9242</td>
<td>.9242</td>
</tr>
<tr>
<td></td>
<td>MX13</td>
<td>.7803</td>
<td>.8333</td>
<td>.7273</td>
</tr>
<tr>
<td></td>
<td>MX14</td>
<td>.6667</td>
<td>.6212</td>
<td>.7121</td>
</tr>
<tr>
<td></td>
<td>MX15</td>
<td>.8712</td>
<td>.8788</td>
<td>.8636</td>
</tr>
<tr>
<td></td>
<td>MX16</td>
<td>.6061</td>
<td>.6667</td>
<td>.5455</td>
</tr>
<tr>
<td></td>
<td>MX17</td>
<td>.7500</td>
<td>.7121</td>
<td>.7879</td>
</tr>
<tr>
<td></td>
<td>MX18</td>
<td>.8864</td>
<td>.8788</td>
<td>.8939</td>
</tr>
<tr>
<td></td>
<td>MX19</td>
<td>.6667</td>
<td>.6212</td>
<td>.7121</td>
</tr>
<tr>
<td></td>
<td>MX20</td>
<td>.8030</td>
<td>.7727</td>
<td>.8333</td>
</tr>
<tr>
<td></td>
<td>MX21</td>
<td>.5152</td>
<td>.4697</td>
<td>.5606</td>
</tr>
<tr>
<td></td>
<td>MX22</td>
<td>.5379</td>
<td>.5758</td>
<td>.5000</td>
</tr>
<tr>
<td></td>
<td>MX23</td>
<td>.5076</td>
<td>.4394</td>
<td>.5758</td>
</tr>
<tr>
<td></td>
<td>MX24</td>
<td>.6364</td>
<td>.7121</td>
<td>.5606</td>
</tr>
<tr>
<td></td>
<td>MX25</td>
<td>.2879</td>
<td>.2273</td>
<td>.3485</td>
</tr>
<tr>
<td></td>
<td>MX26</td>
<td>.5530</td>
<td>.5606</td>
<td>.5455</td>
</tr>
<tr>
<td></td>
<td>MX27</td>
<td>.2955</td>
<td>.2424</td>
<td>.3485</td>
</tr>
<tr>
<td></td>
<td>MX28</td>
<td>.3106</td>
<td>.3636</td>
<td>.2576</td>
</tr>
<tr>
<td>AAT Subscale</td>
<td>Item</td>
<td>Total</td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>-------------</td>
<td>------</td>
<td>--------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>WA</td>
<td>WA1</td>
<td>.9924</td>
<td>.9848</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>WA2</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>WA3</td>
<td>.9848</td>
<td>.9848</td>
<td>.9848</td>
</tr>
<tr>
<td></td>
<td>WA4</td>
<td>.9621</td>
<td>.9848</td>
<td>.9394</td>
</tr>
<tr>
<td></td>
<td>WA5</td>
<td>.7348</td>
<td>.7576</td>
<td>.7121</td>
</tr>
<tr>
<td></td>
<td>WA6</td>
<td>.8106</td>
<td>.7879</td>
<td>.8333</td>
</tr>
<tr>
<td></td>
<td>WA7</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>WA8</td>
<td>.9621</td>
<td>.9394</td>
<td>.9848</td>
</tr>
<tr>
<td></td>
<td>WA9</td>
<td>.9394</td>
<td>.9697</td>
<td>.9091</td>
</tr>
<tr>
<td></td>
<td>WA10</td>
<td>.6970</td>
<td>.6970</td>
<td>.6970</td>
</tr>
<tr>
<td></td>
<td>WA11</td>
<td>.9470</td>
<td>.9242</td>
<td>.9697</td>
</tr>
<tr>
<td></td>
<td>WA12</td>
<td>.9167</td>
<td>.9394</td>
<td>.8939</td>
</tr>
<tr>
<td></td>
<td>WA13</td>
<td>.4318</td>
<td>.3788</td>
<td>.4848</td>
</tr>
<tr>
<td></td>
<td>WA14</td>
<td>.8182</td>
<td>.8182</td>
<td>.8182</td>
</tr>
<tr>
<td></td>
<td>WA15</td>
<td>.6970</td>
<td>.7121</td>
<td>.6818</td>
</tr>
<tr>
<td></td>
<td>WA16</td>
<td>.8561</td>
<td>.8636</td>
<td>.8485</td>
</tr>
<tr>
<td></td>
<td>WA17</td>
<td>.6742</td>
<td>.6212</td>
<td>.7273</td>
</tr>
<tr>
<td></td>
<td>WA18</td>
<td>.9242</td>
<td>.9394</td>
<td>.9091</td>
</tr>
<tr>
<td></td>
<td>WA19</td>
<td>.3788</td>
<td>.4848</td>
<td>.2727</td>
</tr>
<tr>
<td></td>
<td>WA20</td>
<td>.8561</td>
<td>.8788</td>
<td>.8333</td>
</tr>
<tr>
<td></td>
<td>WA21</td>
<td>.4470</td>
<td>.5909</td>
<td>.3030</td>
</tr>
<tr>
<td></td>
<td>WA22</td>
<td>.8182</td>
<td>.8939</td>
<td>.7424</td>
</tr>
<tr>
<td></td>
<td>WA23</td>
<td>.3636</td>
<td>.4091</td>
<td>.3182</td>
</tr>
<tr>
<td></td>
<td>WA24</td>
<td>.5379</td>
<td>.4848</td>
<td>.5909</td>
</tr>
<tr>
<td></td>
<td>WA25</td>
<td>.4091</td>
<td>.3636</td>
<td>.4545</td>
</tr>
<tr>
<td></td>
<td>WA26</td>
<td>.4015</td>
<td>.4545</td>
<td>.3485</td>
</tr>
<tr>
<td></td>
<td>WA27</td>
<td>.4470</td>
<td>.4697</td>
<td>.4242</td>
</tr>
<tr>
<td></td>
<td>WA28</td>
<td>.2273</td>
<td>.2424</td>
<td>.2121</td>
</tr>
<tr>
<td></td>
<td>WA29</td>
<td>.5758</td>
<td>.5758</td>
<td>.5758</td>
</tr>
<tr>
<td></td>
<td>WA30</td>
<td>.2727</td>
<td>.2424</td>
<td>.3030</td>
</tr>
</tbody>
</table>
Item Gender Bias Analysis

All of the items on both the CFAT and AAT were examined for the presence of
gender bias using the WLS logistic regression method outlined in Zumbo (1999) and
endorsed by Swaminatham and Rogers (1990). The results of the DIF analyses for the
CFAT and AAT items are presented in Tables 6 and 7, respectively. The only item that
met the first (required) criterion for DIF was item WA21 on the AAT. The Chi-squared
(2-df) p-value was 0.002, which exceeds the criterion recommended by Zumbo (1999;
Zumbo & Hubley, 1998). The second step, which helps determine whether the DIF is
meaningful or not (in terms of the effect size), also helps determine which type of DIF is
displayed: Uniform or Non-uniform.

The results of the secondary analysis (per Zumbo, 1999) are presented in Table 8.
The R² values were obtained from the logistic regression analyses already performed.
The results of this analysis indicate that the item is displaying uniform DIF, of sufficient
magnitude that further examination of its psychometric properties are warranted.

Distractor Analyses

The distractor analyses were performed separately for the male and female
participant groups. With some exceptions, the correct answers were the most frequently
selected. Typically, there was either one dominant distractor or three distractors being
selected equally often, but neither pattern appears consistently. The distractor analyses
results for the CFAT and AAT items are presented in Tables 9 and 10, respectively.

Several issues arose that necessitate comment. First, the correct response for Item
VS15 on the Verbal Skills subscale of the CFAT was endorsed by only 41.4% of the
respondents. Two incorrect responses were endorsed 34.4% and 20.3% of the time. This
pattern seems to indicate that the participants were rushed for time and filling in random responses. This random response

Table 6

### P-values for the 2df Chi-squared DIF Analysis for Females versus Males on the CFAT Items

<table>
<thead>
<tr>
<th>CFAT Subscale</th>
<th>Item</th>
<th>DIF p-value</th>
<th>DIF?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Chi-squared 2-df</td>
<td></td>
</tr>
<tr>
<td>Verbal Skills</td>
<td>VS1</td>
<td>.884</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>VS2</td>
<td>.506</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>VS3</td>
<td>.523</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>VS4</td>
<td>.803</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>VS5</td>
<td>.994</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>VS6</td>
<td>.634</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>VS7</td>
<td>.037</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>VS8</td>
<td>.907</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>VS9</td>
<td>.934</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>VS10</td>
<td>.070</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>VS11</td>
<td>.633</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>VS12</td>
<td>.926</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>VS13</td>
<td>.842</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>VS14</td>
<td>.147</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>VS15</td>
<td>.155</td>
<td>NO</td>
</tr>
<tr>
<td>Spatial Awareness</td>
<td>SA1</td>
<td>.065</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>SA2</td>
<td>.503</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>SA3</td>
<td>.712</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>SA4</td>
<td>.897</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>SA5</td>
<td>.687</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>SA6</td>
<td>.911</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>SA7</td>
<td>.263</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>SA8</td>
<td>.302</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>SA9</td>
<td>.796</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>SA10</td>
<td>.107</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>SA11</td>
<td>.892</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>SA12</td>
<td>.853</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>SA13</td>
<td>.666</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>SA14</td>
<td>.316</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>SA15</td>
<td>.499</td>
<td>NO</td>
</tr>
</tbody>
</table>
Table 6 continued

<table>
<thead>
<tr>
<th>CFAT Subscale</th>
<th>Item</th>
<th>DIF p-value</th>
<th>DIF?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Chi-squared 2-df</td>
<td></td>
</tr>
<tr>
<td>PS</td>
<td>PS1</td>
<td>.122</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS2</td>
<td>.644</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS3</td>
<td>.376</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS4</td>
<td>.633</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS5</td>
<td>.948</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS6</td>
<td>.606</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS7</td>
<td>.151</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS8</td>
<td>.842</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS9</td>
<td>.243</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS10</td>
<td>.637</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS11</td>
<td>.535</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS12</td>
<td>.673</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS13</td>
<td>.318</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS14</td>
<td>.953</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS15</td>
<td>.555</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS16</td>
<td>.094</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS17</td>
<td>.800</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS18</td>
<td>.474</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS19</td>
<td>.202</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS20</td>
<td>.883</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS21</td>
<td>.104</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS22</td>
<td>.356</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS23</td>
<td>.803</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS24</td>
<td>.585</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS25</td>
<td>.470</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS26</td>
<td>.254</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS27</td>
<td>.141</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS28</td>
<td>.631</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS29</td>
<td>.581</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>PS30</td>
<td>.714</td>
<td>NO</td>
</tr>
</tbody>
</table>
Table 7

P-values for the 2df Chi-squared DIF Analysis
for Females versus Males on the AAT Items

<table>
<thead>
<tr>
<th>AAT Subscale</th>
<th>Item</th>
<th>Chi-squared 2-df DIF p-value</th>
<th>DIF?</th>
</tr>
</thead>
<tbody>
<tr>
<td>MX</td>
<td>MX1</td>
<td>299</td>
<td>NO</td>
</tr>
<tr>
<td>MX2</td>
<td>.478</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX3</td>
<td>.425</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX4</td>
<td>.690</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX5</td>
<td>.011</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX6</td>
<td>.707</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX7</td>
<td>.989</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX8</td>
<td>.875</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX9</td>
<td>.813</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX10</td>
<td>.262</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX11</td>
<td>.247</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX12</td>
<td>.960</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX13</td>
<td>.273</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX14</td>
<td>.406</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX15</td>
<td>.433</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX16</td>
<td>.342</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX17</td>
<td>.403</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX18</td>
<td>.923</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX19</td>
<td>.344</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX20</td>
<td>.504</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX21</td>
<td>.315</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX22</td>
<td>.244</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX23</td>
<td>.156</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX24</td>
<td>.117</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX25</td>
<td>.166</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX26</td>
<td>.777</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX27</td>
<td>.235</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>MX28</td>
<td>.256</td>
<td>NO</td>
<td></td>
</tr>
</tbody>
</table>
Table 7 continued

<table>
<thead>
<tr>
<th>AAT Subscale</th>
<th>Item</th>
<th>DIF p-value</th>
<th>DIF?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Chi-squared 2-df</td>
<td></td>
</tr>
<tr>
<td>WA</td>
<td>WA1</td>
<td>.401</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA2</td>
<td>n/a</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA3</td>
<td>.230</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA4</td>
<td>.448</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA5</td>
<td>.840</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA6</td>
<td>.055</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA7</td>
<td>n/a</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA8</td>
<td>.231</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA9</td>
<td>.392</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA10</td>
<td>.999</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA11</td>
<td>.191</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA12</td>
<td>.558</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA13</td>
<td>.367</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA14</td>
<td>.972</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA15</td>
<td>.689</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA16</td>
<td>.889</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA17</td>
<td>.078</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA18</td>
<td>.178</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA19</td>
<td>.020</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA20</td>
<td>.699</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA21</td>
<td>.002</td>
<td>Possibly</td>
</tr>
<tr>
<td></td>
<td>WA22</td>
<td>.042</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA23</td>
<td>.418</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA24</td>
<td>.426</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA25</td>
<td>.402</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA26</td>
<td>.393</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA27</td>
<td>.657</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA28</td>
<td>.306</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA29</td>
<td>.268</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>WA30</td>
<td>.673</td>
<td>NO</td>
</tr>
</tbody>
</table>

*n/a values: These values cannot be calculated because there was no variance in the item response patterns.*
Table 8

WLS R-squared Calculations for Uniform and Non-Uniform DIF (Zumbo, 1999)

<table>
<thead>
<tr>
<th>Step of Analysis</th>
<th>WLS R-squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>.114</td>
</tr>
<tr>
<td>Step 2</td>
<td>.294</td>
</tr>
<tr>
<td>Step 3</td>
<td>.313</td>
</tr>
<tr>
<td>Step 3 – Step 1</td>
<td>.199 *</td>
</tr>
<tr>
<td>Step 3 – Step 2</td>
<td>.019 **</td>
</tr>
</tbody>
</table>

* this value indicates that the DIF is meaningful in magnitude
** this value indicates that the DIF is uniform in nature

pattern is expected in later items on a timed test where there is no penalty for guessing. A similar pattern is seen on the WA subscale Item WA30.

A similar response pattern is seen in the responses to Spatial Awareness Items SA12 through SA 15 inclusive. Since the largest proportion of respondents endorsed the correct response, and the rest of the responses are spread widely among the other choices, we can assume that this is another case of random guessing on later, more difficult questions.

A different situation arises on the WA subscale Item WA13. Here, the correct response was endorsed less frequently than an incorrect response. Thorough examination of this question is warranted, to determine why the participants chose the wrong answer
almost 50% of the time. Possible problems include poor wording of the question and/or the responses\textsuperscript{33}. Similar problems are seen with WA subscale Items WA23 and WA28.

Another situation arises with the responses to WA subscale Item WA19. The correct response was endorsed less than 50% of the time (38.2%), while an incorrect response was endorsed a large proportion of the time (30.5%). This indicates that that particular incorrect response may be a distractor for the respondents. Similar problems are seen with the WA subscale Items WA21, WA25, WA26 and WA27.

**TABLE 9 AND TABLE 10**

Not included in IMTA proceedings. For further information contact Directorate Human Resource Research and Evaluation – 2 (613 996 0308)

**Test Concordance**

The form of concordance study, or test equating, utilized has been labeled by Crocker & Algina (1986) as *Design ‘B’ Linear Equating*. Within this framework, all participants in the study complete both instruments. Two groups are formed in order to account for possible order effect.

The linear equation for transforming the raw score of the second test to the raw score of the first test is:

\[ Y^* = a(X - c) + d \]  \hspace{1cm} \text{(Equation 1)}

where ‘\(Y^*\)’ is the predicted score on the CFAT, ‘\(X\)’ is the observed score on the AAT, and ‘\(a\)’, ‘\(c\)’ and ‘\(d\)’ are constants calculated from the samples obtained in the concordance study. This is a simple regression equation that uses the observed sample means and

\text{\textsuperscript{33} It should be noted that, due to the confidential nature of the CFAT and AAT tests, this researcher did not have access to the raw questions and answers to provide more thorough analysis of the possible problems identified here.}
variances to determine the constants that define the relationship. See Crocker and Algina (1986) for a more detailed explanation of this relationship.

The results of the concordance study revealed that the values for the constants that fit this data set are ‘a’ = 1.033204, ‘c’ = 41.0175, and ‘d’ = 43.8295. Thus, based on the sample obtained, Equation 1 becomes:

\[ Y^* = 1.033204(X - 41.0175) + 43.8295 \]  

(Equation 2)

where ‘\(Y^*\)’ is the predicted score on the CFAT and X is the observed score on the AAT. Presented in Table 11 is the conversion table for converting scores on the AAT to scores on the CFAT based on the results of this study.

CONCLUSIONS

The results of this study provide support for the use of the AAT as a proxy for the CFAT in times of need. The demographics seem to indicate that the participants responded in a similar manner to those reported by Zumbo & Hubley (1998). That being said, care should be taken in that several issues with both the CFAT and AAT have been identified that bear further examination.

The reliabilities for two of the three CFAT subscales and the AAT subscales were acceptably high. There was some question, however, about the observed reliability of the CFAT Verbal Skills subscale. In addition, the item difficulties for the CFAT Verbal Skills subscale did not seem to model the same pattern as the other two CFAT subscales. No information was available for the AAT scales.

While there did not appear to be any evidence of gender bias in the CFAT items, there was one item on the AAT that required further examination. In addition, no test for other forms
of bias was performed. There may well be other bias issues to be addressed with one or the other of these tests. The distractor analyses also revealed some issues that bear further exploration.

Finally, the issue of Test Concordance. It should be noted that it is possible to produce a conversion table for any two (or more) tests; the crux of the issue is whether the tests are reliably assessing the same constructs. This study has produced a reliable concordance, or conversion, table using an equipercentile equating methodology. However, no data has been collected, nor has any evidence been provided, that allows us to draw any conclusions vis-à-vis the validity of equating these two scales.

Table 11

<table>
<thead>
<tr>
<th>AAT Score</th>
<th>Y*</th>
<th>CFAT score</th>
<th>Raw Score</th>
<th>Y*</th>
<th>CFAT score</th>
</tr>
</thead>
<tbody>
<tr>
<td>58</td>
<td>61.37589</td>
<td>60</td>
<td>29</td>
<td>31.41297</td>
<td>31</td>
</tr>
<tr>
<td>57</td>
<td>60.34268</td>
<td>60</td>
<td>28</td>
<td>30.37977</td>
<td>30</td>
</tr>
<tr>
<td>56</td>
<td>59.30948</td>
<td>59</td>
<td>27</td>
<td>29.34656</td>
<td>29</td>
</tr>
<tr>
<td>55</td>
<td>58.27628</td>
<td>58</td>
<td>26</td>
<td>28.31336</td>
<td>28</td>
</tr>
<tr>
<td>54</td>
<td>57.24307</td>
<td>57</td>
<td>25</td>
<td>27.28015</td>
<td>27</td>
</tr>
<tr>
<td>53</td>
<td>56.20987</td>
<td>56</td>
<td>24</td>
<td>26.24695</td>
<td>26</td>
</tr>
<tr>
<td>52</td>
<td>55.17666</td>
<td>5</td>
<td>23</td>
<td>25.21375</td>
<td>25</td>
</tr>
<tr>
<td>51</td>
<td>54.14346</td>
<td>54</td>
<td>22</td>
<td>24.18054</td>
<td>24</td>
</tr>
<tr>
<td>50</td>
<td>53.11026</td>
<td>53</td>
<td>21</td>
<td>23.14734</td>
<td>23</td>
</tr>
<tr>
<td>49</td>
<td>52.07705</td>
<td>52</td>
<td>20</td>
<td>22.11413</td>
<td>22</td>
</tr>
<tr>
<td>48</td>
<td>51.04385</td>
<td>51</td>
<td>19</td>
<td>21.08093</td>
<td>21</td>
</tr>
<tr>
<td>47</td>
<td>50.01064</td>
<td>50</td>
<td>18</td>
<td>20.04773</td>
<td>20</td>
</tr>
<tr>
<td>46</td>
<td>48.97744</td>
<td>49</td>
<td>17</td>
<td>19.01452</td>
<td>19</td>
</tr>
<tr>
<td>45</td>
<td>47.94424</td>
<td>48</td>
<td>16</td>
<td>17.98132</td>
<td>18</td>
</tr>
<tr>
<td>44</td>
<td>46.91103</td>
<td>47</td>
<td>15</td>
<td>16.94811</td>
<td>17</td>
</tr>
<tr>
<td>43</td>
<td>45.87783</td>
<td>46</td>
<td>14</td>
<td>15.91491</td>
<td>16</td>
</tr>
<tr>
<td>42</td>
<td>44.84462</td>
<td>45</td>
<td>13</td>
<td>14.88171</td>
<td>15</td>
</tr>
<tr>
<td>41</td>
<td>43.81142</td>
<td>44</td>
<td>12</td>
<td>13.8485</td>
<td>14</td>
</tr>
<tr>
<td>40</td>
<td>42.77821</td>
<td>43</td>
<td>11</td>
<td>12.8153</td>
<td>13</td>
</tr>
<tr>
<td>39</td>
<td>41.74501</td>
<td>42</td>
<td>10</td>
<td>11.78209</td>
<td>12</td>
</tr>
<tr>
<td>38</td>
<td>40.71181</td>
<td>41</td>
<td>9</td>
<td>10.74889</td>
<td>11</td>
</tr>
<tr>
<td>37</td>
<td>39.6786</td>
<td>40</td>
<td>8</td>
<td>9.715685</td>
<td>10</td>
</tr>
<tr>
<td>36</td>
<td>38.6454</td>
<td>39</td>
<td>7</td>
<td>8.682481</td>
<td>9</td>
</tr>
<tr>
<td>35</td>
<td>37.61219</td>
<td>38</td>
<td>6</td>
<td>7.649277</td>
<td>8</td>
</tr>
<tr>
<td>34</td>
<td>36.57899</td>
<td>37</td>
<td>5</td>
<td>6.616073</td>
<td>7</td>
</tr>
<tr>
<td>33</td>
<td>35.54579</td>
<td>36</td>
<td>4</td>
<td>5.582868</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td>---</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>32</td>
<td>34.51258</td>
<td>35</td>
<td>3</td>
<td>4.549664</td>
<td>5</td>
</tr>
<tr>
<td>31</td>
<td>33.47938</td>
<td>33</td>
<td>2</td>
<td>3.51646</td>
<td>4</td>
</tr>
<tr>
<td>30</td>
<td>32.44617</td>
<td>32</td>
<td>1</td>
<td>2.483256</td>
<td>2</td>
</tr>
</tbody>
</table>
References


ORGANIZATIONAL, SOCIAL AND DEMOGRAPHIC CHANGE
IN THE CANADIAN FORCES: 1976 TO 2001

T. Wait
Directorate Of Strategic Human Resources
Department of National Defence
October 2002

This research note is one in a series of papers concerning Canadian trends and challenges internal and external to the Canadian Forces, and is part of the HR2020 Project. The document provides a brief summary of organizational, social, and demographic change in the Canadian Regular Force from 1976 to 2001. Changes during this thirty-year time period to the structure of the Canadian Forces (CF), to the profile of enrollees and CF members, and to CF member’s families, are explored. Modifications to the composition of the CF are evaluated against shifts in similar areas in the Canadian population. For the most part, changes in the CF are equivalent to those in the Canadian labour force. The implications of these organizational, social, and demographic changes for the CF are considered.

Résumé
Introduction
1. The primary focus of this paper is the exploration of organizational, social, and demographic changes and trends in the Regular Force of the Canadian Forces (CF) from 1976 to 2001. To better understand some of these changes, findings from the CF are compared with shifts in the Canadian population over this same time span. To understand how these organizational, social, and demographic changes might affect the CF, their implications are identified and discussed.

2. As with the Canadian labour force, the composition of the Canadian Forces (CF) has evolved over the past 25 years. This has largely been the result of socio-demographic, technological and economic trends, as well as downsizing initiatives taken in the 1990s. While the general organizational structure and design of the CF has not changed much in the past 25 years, the demographic profile of serving members and their families has changed considerably.

Canadian Forces Regular Strength
3. By the end of 2001, the Canadian Forces (CF) regular force membership had declined to 58,481, a decrease of 3,159 from the 1996 total of 61,640 (see Figure 1). There has been a continual decline in regular force personnel since 1986, with the Force Reduction Program (FRP) leading to the release of approximately 14,000 personnel as a means to cut defence related costs. The unexpected outcome of such a large reduction in personnel, accompanied by a virtual elimination of recruitment of NCMs, was an inability to exercise proper succession planning in a number of military occupations. This has resulted in an attrition rate significantly higher than the intake rate (see Figure 2).

FIGURE 1: CANADIAN FORCES REGULAR STRENGTH, 1976-2001

4. In an effort to remedy the current “crisis” the CF has been strengthening their recruitment effort, and will continue with this effort for the next three years. The recruitment surge began in 1999/2000 with the successful recruitment of 2,623 individuals. 2000/2001 saw the recruitment of 3,220 new members. While the CF is aiming to recruit 7,000 new members during the 01/02 fiscal year, by end of March 02, it had successfully recruited 5,404. The program aims to recruit 6,122 during 02/03, and 6,025 during 03/04. The CF will then hope to stabilize its annual recruitment at 5,000. Under the current environment (low attrition, questionable economy) the CF could recover from its recruitment/retention crisis fairly soon, however, if attrition reverts to its higher levels (approximately seven percent attrition per year), the CF will likely recover
by 2006/2007. It is likely that the current surge in recruiting will lead to an increase in releases. Data show that the CF loses approximately 33% of all new (NCM) recruits within 3 years. With approximately 6,000 new recruits, the CF can expect to lose 2,000 or so over the next 3 years, or 700 per year for each recruit cohort.
5. **Figure 2: Attrition and Intake Comparison,**

![Chart showing attrition and intake comparison]

*The CF Rank Structure*

6. The division of officers and non-commissioned members (NCMs) has remained relatively stable throughout the last 25 years. While total strength has declined slightly since 1996, the decline is evident in the ranks of junior NCMs, who accounted for 60% of total strength in 1976, and were down to 56% in 2001. Senior NCMs have remained stable at 21%, while both junior and senior officer groups have increased from 13% and 5% in 1976, to 15% and 7.3% in 2001, respectively (see Figures 3 and 4).

7. Prior to 1996, the FRP limited the number of new recruits, resulting in a dramatic decline in the number of Privates and Corporals from 46% in 1976 to 41% in 1996. However, recruitment has increased since 1996 resulting in an increase in Privates from 8% in 1996, to 11.5% in 2001. Given the decline in Privates until 1996, the CF continues to see a decline in Corporals and Master Corporals right through to 2001, as fewer Privates are available for promotion. There has been a similar decline in the number of Captains (since 1996) as recruitment was lean during those years (see Figure 5).
FIGURE 3: CF STRENGTH BY RANK, 1976-2001
FIGURE: 4 CF REGULAR FORCE RANK 1976-2001
Canadian Forces Recruitment 1990-2000

<table>
<thead>
<tr>
<th>INTAKE YEAR</th>
<th>MALES (% of Total)</th>
<th>FEMALES (% of Total)</th>
<th>TOTAL ENROLLED</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990/91</td>
<td>4,646 (83%)</td>
<td>884 (16%)</td>
<td>5,610</td>
</tr>
<tr>
<td>1991/92</td>
<td>2,188 (85%)</td>
<td>383 (15%)</td>
<td>2,571</td>
</tr>
<tr>
<td>1992/93</td>
<td>1,308 (87%)</td>
<td>193 (13%)</td>
<td>1,501</td>
</tr>
<tr>
<td>1993/94</td>
<td>1,789 (89%)</td>
<td>221 (11%)</td>
<td>2,011</td>
</tr>
<tr>
<td>1994/95</td>
<td>1,629 (92%)</td>
<td>147 (8%)</td>
<td>1,776</td>
</tr>
<tr>
<td>1995/96</td>
<td>1,478 (89%)</td>
<td>163 (10%)</td>
<td>1,656</td>
</tr>
<tr>
<td>1996/97</td>
<td>3,391 (88%)</td>
<td>458 (12%)</td>
<td>3,849</td>
</tr>
<tr>
<td>1997/98</td>
<td>1,533 (83%)</td>
<td>303 (17%)</td>
<td>1,840</td>
</tr>
<tr>
<td>1998/99</td>
<td>1,069 (83%)</td>
<td>214 (17%)</td>
<td>1,286</td>
</tr>
<tr>
<td>1999/2000</td>
<td>1,788 (87%)</td>
<td>273 (13%)</td>
<td>2,064</td>
</tr>
</tbody>
</table>

8. Since 1996, the number of highest ranking officers has declined slightly from 78 down to 70 Generals. While during 1996 the ratio of Generals to CF members was 1:790, the ratio is currently 1 General per 835 members. The years between 1996 through to 2001 have seen a slight decline in the number of Majors, while the number of Colonels and Lieutenant Colonels have seen a minor increase. With the current push to recruit, the CF should expect to see a greater number of junior ranks entering the system, thus swelling the ranks of Corporal/Master Corporal and Lieutenant/Captain in the coming years.

9. The military occupation structure (MOS) is currently undergoing a formal review. The Military Occupational Structure Analysis, Redesign and Tailoring (MOSART) Project will provide a multi-level assessment of the current MOS policies and structural framework, investigating whether the current MOS supports the CF mission and vision to the best advantage or, if not, how it might be changed to do so more effectively.

10. In 1986 there were 100 NCM and 42 officer military occupation classifications. However, since that time the CF has seen significant downsizing and alterations to the MOC structure. Currently there are 73 NCM and 32 officer MOCs.

New Enrollees in 2001

11. There were 4,001 new enrollees in the Canadian regular forces during the 2001 calendar year. Eighty-three percent (3,291) of them were male, 18% (709) were female. Prior to 1973, only about 5% of enrollees were female, this increased to 10% in 1973 and was as high as 21% in 1992.

12. Seventy-seven percent of enrollees in 2001 were English speaking, while the remaining 23% were French speaking.

13. In keeping with the Canadian population in general, new enrollees continue to age. The majority (67%) of new enrollees in 2001 was between the age of 17 to 24, 26% were between 25 and 34, 7% were between 35 and 44, and the remaining new members (0.3%) were over the age of 45. In 1996, the average age of enrollees was 22.5 and 50% of the enrollees were under 21 years of age.

14. Reflecting patterns of increased education in Canada, enrollee’s education attainment continues to increase. The largest proportion of new enrollees in 2001, had a high school (or equivalent) diploma (47%). Remarkably, slightly more than 21% of our new regular
force members have less than a high school education (meaning some but not all of a secondary education). However, in previous years many enrollees had achieved only some elementary education. Thirty-one percent of new enrolees in 2001 have a post secondary education ranging from college right through to doctoral degrees.

15. The greatest portion of new regular force members in 2001 were single (76%), however, 21% were either married or living with a common-law partner. The remaining 3% were separated, divorced or widowed. Eighty-nine percent of 2001 enrollees had no child dependents, however the remaining 11% had anywhere from 1 to 5 children dependent on them. The majority of new enrollees (59%) joined the army, 25% joined the air force, and 17% joined the navy.

*Education of New Members by Sex, 2001*

16. During the 2001 calendar year 709 females entered the regular Canadian Forces. The greatest number of the new female members (41%) had a high school degree or equivalent. Twenty-four percent of new female recruits had a university education (including both undergraduate and graduate/post graduate degrees). Another 22% of new female CF members had a post secondary education, while the remaining 13% had less that a high school degree.

17. During the same year 3,291 males were enrolled as regular force members in the CF. Forty-nine percent of the new enrollees had a high school degree or equivalent. Fourteen percent of male enrollees had a post secondary education, and another 14% had a university education (including both undergraduate and graduate/post graduate degrees). Finally, a full 23% of new male enrollees had less than a high school education (see Figure 6).

*Figure 6: Education Level of New Enrollees by Sex*

18. The majority of new male enrollees is within the 17 to 24 year age cohort, and account for over 71% of all new male CF members. The next largest age group is those between the ages of 25 and 34, who account for another 24% of the total. In total, 95% of all male enrollees are under the age of 35. The new female CF members are on average older than their male counterparts. Only 44% of new female members are between 17 and 24 years of age. Another 36% are between the ages of 25 and 34. However, surprisingly, the remaining group (those over the age of 35) is relatively large, accounting for almost 19% of new female CF enrollees (see Figure 7).
19. The majority of new enrollees, both male and female, have joined the CF at entry level positions. Sixty-four percent of new female members entered the CF as privates, and another 16% entered as officer cadets. Similarly, 79% of new male enrollees entered the CF as privates, while another 10% entered as officer cadets. However, interestingly a full 7% of female enrollees entered the CF at levels of lieutenant or higher and another 12% joined the forces at the corporal level. Four percent of new male enrollees entered the forces at the lieutenant level or higher, and another 6% entered the CF at the corporal level or higher.

20. Not surprisingly, those who entered at higher ranks were on average older than those who joined the CF at the entry level. Twenty-four percent of male officer enrollees who entered at higher than entry level were over the age of 35. Similarly, 26 percent of female officer enrollees who entered at higher than entry level were over the age of 35.

Family Status of New Members by Sex, 2001

21. Fifty-nine percent of new female members were single, another 29% were either married or living in a common-law relationship. Over 5 percent of the new female entrants were married or living common-law with another service member. The remaining 7% were separated, divorced or widowed.

22. Of the single female entrants, 9 percent had dependant children, and were presumably single parents. In addition, of those new female entrants who were either separated or divorced, 62% had dependant children.

23. Eighty percent of the new male CF members were single at the time of enrollment. Another 19% were either married or living in a common-law relationship. Less than one half of one percent (0.3%) were either married or living common-law with another service member. Over 90% of the new male entrants had no dependant children at the time of entry, however, of the single male entrants almost 2% had dependant children, the majority (83%) of which were under the age of 7.

Country of Birth

24. The greatest proportion of new entrants was born in Canada, 91% of male and female entrants. Germany accounted for a full 1 percent of each of the male and female entrants. While the CF has recently conducted a full census of its personnel, the diversity data have yet to be released, therefore there is little in the way of published information pertaining to the ethnicity of new entrants. Having said that, the CF has recently begun to collect diversity data concerning potential applicants through a contact survey.
Diversity of Those Attracted to the CF
25. The Canadian Forces Contact Survey Annual Monitoring Report, June 2000-June 2001, identifies attraction influences, advertising factors and demographic characteristics of interested people who come to Canadian Forces Recruiting Centres. In keeping with the demographic pool from which the CF has traditionally recruited, the “typical contact” is a “white, English-speaking male, between 14-19 years of age, with no previous military experience, with an educational level somewhere between grade 10 and grade 12, working part-time, likely attending school full-time and who currently resides in a city”.
26. Eighteen percent of contacts surveyed identified themselves as a member of a visible minority group, while 7% identified themselves as aboriginal (see Figure 8).
27. **Figure 8: Diversity of CF Contacts, 2000/2001**

<table>
<thead>
<tr>
<th>ABORIGINAL:</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not an aboriginal person</td>
<td>387</td>
<td>3.1%</td>
</tr>
<tr>
<td>First Nation</td>
<td>191</td>
<td>.6%</td>
</tr>
<tr>
<td>Metis</td>
<td>81</td>
<td>.9%</td>
</tr>
<tr>
<td>Inuit</td>
<td>13</td>
<td>.3%</td>
</tr>
<tr>
<td>Total</td>
<td>416</td>
<td>100%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VISIBLE MINORITY:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Not a Visible Minority</td>
<td>278</td>
<td>1.6%</td>
</tr>
<tr>
<td>Yes, I am a Visible Minority</td>
<td>626</td>
<td>8.4%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VISIBLE MINORITY GROUP:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>114</td>
<td>8.2%</td>
</tr>
<tr>
<td>Chinese</td>
<td>99</td>
<td>5.8%</td>
</tr>
<tr>
<td>South Asian/East Indian</td>
<td>89</td>
<td>4.2%</td>
</tr>
<tr>
<td>Mixed race or colour</td>
<td>87</td>
<td>3.9%</td>
</tr>
<tr>
<td>Non-White West Asian</td>
<td>70</td>
<td>1.2%</td>
</tr>
<tr>
<td>Non-White Latin American</td>
<td>60</td>
<td>.6%</td>
</tr>
<tr>
<td>Filipino</td>
<td>54</td>
<td>.6%</td>
</tr>
<tr>
<td>Korean</td>
<td>32</td>
<td>.1%</td>
</tr>
<tr>
<td>Southeast Asian</td>
<td>17</td>
<td>.7%</td>
</tr>
<tr>
<td>Japanese</td>
<td>4</td>
<td>.6%</td>
</tr>
<tr>
<td>Total</td>
<td>626</td>
<td>100%</td>
</tr>
</tbody>
</table>
The CF Regular Force Population

Age of CF Members

28. In keeping with the Canadian population and the Canadian labour force, the CF continues to age. At the end of 1996, the largest group of CF members was in the 30-34 age cohort. In 2001, the largest group of CF members were between the ages of 35 and 39, with the mean age of CF personnel being 35 years old (see Figure 9). There has also been a large increase in the proportion of those between the ages of 40 and 44, as well as increases in those 45-49 and 50-54.

29. The average age of senior officers (rank of Major to General) is 43.2 years old, while the average age of junior officers (OCDT to Captain) is 32.3. The average age of senior NCMs is 41.4 years and the average age of junior NCMs is 32 years old.

Figure 9: CF Regular Force Age Profile

Education of CF Members

30. Educational attainment in Canada has continued to increase for several decades. In fact, the percentage of those 15 years and older, without a degree, certificate or diploma has declined from 48% in 1986 to 37% in 1996. The education level of regular force CF members continues to increase, following the Canadian trend. In 2001, 13% of CF personnel possessed a university degree (undergraduate, masters and doctorate), changing little since 1996. Eighteen percent of CF members had some form of post secondary education (including technical and college training), up almost 8% since 1996. Thirty seven percent of regular CF members possessed a high school degree, while 32% had achieved less that a high school diploma (see Figure 10).
31. **Figure 10: Regular Force Education, 2001**

32. Of those regular CF members who reported having achieved less than a high school diploma, 54% were over the age of 35, suggesting that they entered the Canadian Forces at a time when the education criteria were not as stringent as they are today. The largest age cohort to report having earned less than a high school (or equivalent) diploma were those between the ages of 35 and 44 (44%). Another 10 percent of CF members reporting the same results were over the age of 45. The remaining 46% of CF personnel reporting less than a high school education were below 35 years of age.

**Women in the CF**

33. Women have increased their presence in virtually all areas of the Canadian Forces. As of 2000, all MOCs except for military Roman Catholic chaplains have been opened to women. Within the army, air, and naval environments, women have been assuming more operational roles. Traditionally, women filled support roles in the CF; however, since the Human Rights Tribunal decision of 1989, the participation of women in operational roles has increased both at the officer and NCM level. With the recent amendment to the Employment Equity Act, the CF is monitoring specific MOCs to improve the participation of women and other designated groups.

34. Between 1996 and 2001, the percentage of women in the CF has increased from 10.7% to 12%, while female representation in the Canadian labour force has remained stable at 46%.

**Figure 11: Female CF Members by Rank (% Regular Force)**
Female representation continues to increase in almost all ranks of the CF. Since 1976 female representation at the senior officer level has gone from zero to 0.05%, female junior officer representation has increased from 1% to approximately 3%, female representation at the senior NCM level has increased from 0.1% to 2%, and finally, female junior NCM representation has increased from 4% to 7% (see Figure 11). If the CF continues to enroll new female members at rates greater than their current representation in the CF (as it has in 2001), then female representation in the CF will continue to rise.

Francophones and Anglophones in the CF

35. The current split between English and French speaking personnel in the Canadian Forces is approximately 3:1 (English to French) at the officer level and 2:1 (English to French) at the NCM level (see Figure 12). French speaking CF regular members account for approximately 27% of the Canadian Forces, down 1% from 1996. This figure slightly exceeds the 1996 Census profile of 23% of the Canadian population reporting French as their mother tongue in 1996 (59% report English as their mother tongue, while 16% report a non-official language as their mother tongue).

**Figure 12: Regular Force Anglophone Francophone Balance, 2001**

36. While there has been a slight rise in Francophone CF members at the officer and senior NCM levels, there has been a consistent decline in French speaking junior NCMs since 1976 (from 60% down to 55% of Francophone members) (see Figure 13). Twenty-
three percent of enrollees in 2001 were French speaking (less than the current percentage of Francophone members in the CF).
37. 
Figure 13: Francophone CF Members by Rank, 1976-2001

Country of Birth

<table>
<thead>
<tr>
<th>Country of Birth</th>
<th>% Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>32%</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>18%</td>
</tr>
<tr>
<td>United States</td>
<td>8%</td>
</tr>
<tr>
<td>France</td>
<td>7.3%</td>
</tr>
<tr>
<td>Netherlands</td>
<td>2.5%</td>
</tr>
<tr>
<td>Poland</td>
<td>2.2%</td>
</tr>
<tr>
<td>Jamaica, Philippines</td>
<td>2.0% ea.</td>
</tr>
<tr>
<td>Belgium, Guyana, Haiti, India, Ireland, Portugal, South Africa</td>
<td>1.0% ea.</td>
</tr>
</tbody>
</table>

Aboriginal and Visible minority Representation in the CF

39. The CF has made efforts to increase Aboriginal and visible minority representation within the ranks. The most recent Employment Equity data available is from March 1997 (see Figure 14). At that time almost 97% of the regular CF members were non-Aboriginal and/or non-visible minority. Almost 2 percent of CF members were reported to be visible minorities and another 1.5% reported as being Aboriginal people. During the year 2001, the CF engaged in another full census of personnel, including employment equity status. The results will be available in late spring 2002.
40.

**FIGURE 15: REPRESENTATION OF EE GROUPS (MARCH 1997)**

<table>
<thead>
<tr>
<th></th>
<th>Regular Force Total</th>
<th>Primary Reserve Total</th>
<th>Total CF Total</th>
<th>Canadian Workforce (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td></td>
</tr>
<tr>
<td>Non-Aboriginal / Non-Visible Minority</td>
<td>61,347 (96.7)</td>
<td>50,322 (96.5)</td>
<td>91,669 (96.6)</td>
<td>88%</td>
</tr>
<tr>
<td>Aboriginal Peoples</td>
<td>969 (1.5)</td>
<td>106 (1.0)</td>
<td>1,275 (1.3)</td>
<td>3%</td>
</tr>
<tr>
<td>Visible Minorities</td>
<td>1,141 (1.8)</td>
<td>863 (2.7)</td>
<td>2,004 (2.1)</td>
<td>9%</td>
</tr>
<tr>
<td>Total</td>
<td>63,457 (100)</td>
<td>31,491 (100)</td>
<td>94,948 (100)</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

**CF Members Families**

_Marital Status of CF Members_

41. The proportion of married members in the CF has declined significantly since 1976. In 2001, 49.2% of members were married, while in 1976, 64.5% of CF personnel were married. However other forms of family status have increased to compensate for this decline. While common-law relationships were not recorded in 1976, in 1996 they accounted for almost 9% of CF families, and in 2001 common-law relationships account for a full 12%. Service couples (those either living common-law or married to another service member) have increased from 2% in 1976 to almost 7% in 2001. In 2001, only 3 percent of CF personnel were divorced, while another 3.4% were formally separated (see figure 15).

**FIGURE 16: MARITAL STATUS OF CF MEMBERS**

_Single-Parent Families_

42. In 2001, 6.1% (919) of single CF members had dependents, the largest component of this group had only one dependent. Of those CF personnel who were separated, divorced or widowed, 67% or 2,500 personnel had dependents, the majority of whom had 2 dependents, a family size slightly larger than the Canadian average. The average census family size in Canada as a whole has declined slightly over time. In 1976, the average size of all families was 3.5, while the average size of a lone-parent family was 2.9. The numbers have declined slightly to an average Canadian family size of 3.0 and lone-parent family size of 2.5 in the year 2000.
Size of CF Families

43. In 2001, over 55% of CF members had dependents, up very slightly from 1996. The proportion of CF members without dependents is declining over time, from 55% in 1986, to 46% in 1996, and now 45% in 2001. The number of CF families with 4 or more children has remained relatively stable since 1996, however the longer-term trend is clearly one of decreasing family size. Currently, the largest proportion of CF families with dependents have only 2 children (28%). This form of family unit has seen a gradual but continuous increase from 21% in 1986, and 26% in 1996 (see Figure 16). The largest group of Canadian families also has, on average, 2 children, be they two partner or lone-parent families.

Age of Dependents

44. The average age of all CF child dependents was 10 years in 2001. While as stated before, most CF families had 2 dependents, some had as many as 8 children. On average the age of the first child was approximately 8 1/2 years old, the second was slightly less than 12, the third was just 14 years old. During the year 2000, the majority of Canadian census families had children under the age of 18. Of those families the largest group had children between the ages of 6 and 14 years.

Implications for the CF

Demographics

45. The changing Canadian demographic profile has already had profound impacts on the CF and its organization. Recruitment and retention policies and practices have had to be transformed as the CF faces legislation around the attraction and hiring of employment equity designated groups (women, aboriginals and visible minorities). The traditional male, white, high school (or less) educated 17-24 year old cohort that the CF has targeted in the past is in decline as new immigrants change the face of Canada. While the absolute size of the 17-24 age cohort will remain stable (at approximately 3.5 million out to 2016), as a proportion of the total population the cohort is declining, and change will occur in the composition of the group.

Demand for Skilled Labour

46. The CF has also been caught up in the labour market competition for the best and brightest candidates, both young and old, as the demand for knowledge workers has...
skyrocketed. The demand for well educated knowledge workers has been such that this group has dictated working conditions, pay and benefits. Competition for these individuals has been fierce, driving the CF to offer special signing bonuses, education, compensation and benefit packages in an effort to attract knowledge workers to its doors. As technology continues to change the nature of work, demand for knowledge workers will increase, forcing organizations to compete even harder for skilled workers. Recruitment and retention strategies will likely include: specialized healthcare services as the population continues to age; an environment supportive of work/life balance; challenging work with up-to-date tools to do the job; and respect and recognition, given that unmet worker expectations lead to attrition.

Organisation Renewal

47. The CF has faced organisation re-design over the years in an effort to reduce costs while maintaining efficiency. Alternative service delivery (ASD) and civilianization of many former military occupations has changed the military organisation significantly. The CF will likely continue to explore and implement these methods as competition for scarce government resources continues to be a reality.

Conclusion

48. The CF faces many challenges in the future, many of which are evident today. The workforce is evolving, competition for new skilled workers is intense, families are changing and worker expectations must be satisfied in order to retain valued employees. The CF has continued to meet the challenges through employment equity programs, recruitment and retention programs, quality of life initiatives and healthcare reform.

49. The CF has mirrored Canadian socio-demographic and employment trends over the last 25 years. The force has become older, more highly educated, more reflective of an evolving Canadian demographic profile, and more tolerant of individual differences. The CF has also begun to understand the importance of a flexible, adaptable organisation that continuously measures its performance to ensure that current and future worker expectations are considered and balanced with organizational needs. As the Canadian context continues to change over time, so too will the CF in order to remain relevant and important to Canadians.

References

3. Department of National Defence, DMHRR Data 2001/10/04
   http://hr.dwan.dnd.ca/dmhrr/engraph/library/psr_e.asp
Military HR Strategy 2020, Facing the People Challenges of the Future:
Consultation Results

T. Wait
Directorate Of Strategic Human Resources
Department of National Defence
October 2002

Abstract
Key to the development of HR2020 was a process of consultation and discussion throughout DND. This process included both focus groups and interviews with key personnel to ensure as wide a range of input as possible into the document. D STRAT HR staff conducted 22 interviews of Senior Management within the CF and DND and four focus groups (Chief Warrant Officers, Land, Air, and Sea) designed to elicit personal opinions on past experiences, current practices and possible futures of the CF.

Résumé
La clef au développement de HR2020 était un procédé de consultation et discussion à travers DND. Ce processus a inclus les groupes de foyer et entrevues les personnels de base pour assurer aussi au loin une gamme d'entrée comme possible dans le document. D STRAT HR a conduit 22 entrevues de haute direction dans les CF et DND et quatre groupes de foyer (Adjuntant-chef,officiers, terre, air, et mer) conçus pour obtenir des avis personnels des expériences antérieures, des pratiques en vigueur et possible avenirs du CF.
Executive Summary

Key to the development of HR2020 was a process of consultation and discussion throughout DND. This process included both focus groups and interviews with key personnel to ensure as wide a range of input as possible into the document. D STRAT HR staff conducted 22 interviews of Senior Management within the Canadian Forces (CF) and the Department of National Defence (DND) and four focus groups (Chief Warrant Officers, Land, Air, and Sea) designed to elicit personal opinions on past experiences, current practices and possible futures of the HR within the CF. What follows is a summary of some of the major findings of these consultations. It bears repeating that these conclusions are the views and opinions of the sample and may not necessarily reflect the actual situation or the views of CF members.

Leadership

Leadership in general, across the organisation, as stronger now than ever before
While leadership has improved, the organisation requires greater HR expertise at leadership levels
Movements in and out of HR leadership functions must be less frequent
The CF organisational structure of the future will be flatter and less hierarchical, leaders must work in a less rigid environment, power will be shared both up and down the chain of command
Emphasis will be on “softer skills” i.e., ability to motivate, work in teams, be ethical, strategic and critical thinking, political astuteness

HR Strategy and Planning

The Department continues to operate tactically rather than strategically, struggling with policy versus service delivery and strategic planning versus tactical responses
There continues to be a lack of flexibility in the HR system, even though flexibility and agility will be essential in the future
Personnel policy must reflect changing values and needs of individuals

Culture

The key to organisational change is culture change
Organisation must be creative, at ease with diversity and respectful of individual rights
The CF of the future must narrow the perceived gap between organisational and societal values

People Focus

Quality of life programs are a good first step towards a people focus
Lack of work/life balance reflects a organisation that does not put people first
To meet future personnel and operational requirements, a greater emphasis on personnel and their needs will be essential

Communications

Internal communications of pertinent information is currently awkward and difficult to obtain
The organisation needs to re-negotiate its relationship with media
Internal consultation will, in some cases, replace orders and will be far more transparent

Compensation and Benefits

The organisation has done well in eliminating the CF/private sector wage gap
In the future, pay will not be tied to rank
Benefits packages will be personalised and more inclusive, caring for both members and their families

Education, Career and Professional Development

There is a feeling that experience is being neglected in favour of academic studies
The CF of the future will be, as a learning institution, a viable alternative to costly educational institutions
Professional development, in the future, should include secondments to universities, colleges, industry and other government departments
The selection and promotion of members will change radically, to include initiatives such as competition for postings

Recruitment and Retention

Greater mobility in and out of the organisation as well as between regular and reserve forces will be required
Periods of engagement must be more flexible to attract youth
Today’s recruits are our future leaders, they must be bright, multilingual individuals - the days of “cannon fodder” are over
A comprehensive recruitment strategy including improved quality of life programs and integrated and personalised compensation and benefits packages will be necessary

CONCLUSION

The need for change was a theme that ran through all interviews and focus groups. This change must be more than superficial. Rather, almost all of those participating in this exercise were of the opinion that real change, spearheaded by strong leadership, is required to ensure the survival of the CF into the future.

A CF that is more diverse both in terms of membership and in the way it operates will be the hallmark of an organisation prepared to face the challenges of 2020. Many people expressed misgivings about the ability of the organisation to rise to these challenges. However, almost all felt that steps have been taken in the right direction and they are waiting to see if the leadership will continue on this path forward.

INTRODUCTION

1. HR2020 was intended as an HR response to Defence Strategy 2020. Building upon the initial work undertaken in People in defence – Beyond 2000, HR2020 presented a strategy for HR planning and management within the CF. Key to the development of HR2020 was a process of consultation and discussion throughout DND. This process included both focus groups and interviews with key personnel to ensure as wide a range of input as possible into the document. What follows is a summary of some of the major findings of these consultations.

METHODOLOGY

2. In March and April 2002, the Directorate of Strategic Human Resources (D STRAT HR) staff conducted 22 interviews of Senior Management within the CF and DND (see
Annex D for list of respondents). The interview format was based on a questionnaire containing eight open-ended questions (see Annex A for a list of the questions). These were designed to elicit personal opinions on past experiences, current practices and possible futures of the CF as related to human resource management and planning.

3. While some interviews lasted the scheduled thirty minutes, due to the level of interest of respondents, many interviews continued for periods of one hour or more. Respondents appeared to appreciate the opportunity to voice personal concerns and opinions and to have input into the HR2020 process.

4. In order to reach more people at various levels in the organisation in a relatively short period of time, four focus groups were also organized. These groups included more than 20 people in total and were conducted with Chief Warrant Officers, as well as Land, Air, and Maritime Staff (See Annex C for the Moderator’s Guide used in focus groups).

THROUGHOUT THE INTERVIEWS AND FOCUS GROUPS COMMON THEMES AND IDEAS PERMEATED THE DISCUSSION. HOWEVER, RESPONDENTS CAME UP WITH MANY UNIQUE AND THOUGHT-PROVOKING IDEAS AND COMMENTS AS WELL. TO ASSIST IN THE CONTENT ANALYSIS OF THIS INFORMATION, THE FEEDBACK RECEIVED WAS CODED UNDER A NUMBER OF GENERAL THEMES OR HEADINGS. IN THE FIRST SECTION OF THIS REPORT, INFORMATION IS PRESENTED ON RESPONDENTS’ THOUGHTS REGARDING CURRENT AND PREVIOUS HR PRACTICES. THE SECOND SECTION THEN DEALS WITH IDEAS ON FUTURE DIRECTIONS AND STRATEGIES PERTAINING TO HR IN THE CF.

CURRENT AND PAST PRACTICES

General

ALMOST ALL RESPONDENTS AGREED THAT THE CF WAS BETTER IN TERMS OF HR PRACTICES NOW THAN AT ANY POINT IN THE PAST. MANY REFERRED TO THE FORCE REDUCTION PROGRAM (FRP) AS AN EXAMPLE OF POOR HR PLANNING AND IMPLEMENTATION THAT HAD AN ENDURING IMPACT ON THE STRUCTURE OF THE CF AND THE RELATIONSHIP BETWEEN THE ORGANISATION AND MEMBERS. HOWEVER, THIS IS NOT TO SAY THAT THINGS ARE PROCEEDING TODAY AS WELL AS THEY SHOULD. IN MOST ASPECTS OF HR MANAGEMENT WITHIN THE CF, RESPONDENTS FOUND DEFICIENCIES RESULTING IN A NEED FOR IMPROVEMENT.

Leadership

5. A common refrain from many respondents was captured by one person, who commented that, “The key strength today is stronger leadership at all levels in the organisation”. What continues to weaken the effectiveness of this leadership is the fact that there is little continuity in senior positions. Frequent personnel changes do not allow leaders to occupy positions long enough to learn the role of an HR manager and make effective changes.

6. At some levels of leadership, some respondents argued there is a lack of HR professionalism and HR experience. People are often appointed to positions without a grounding in HR. One other criticism directed more towards the responsibilities of leaders rather than the leadership itself, was that those in charge often have too much to handle. “The span of control of the COS (Chief of Staff) is too broad. He has all the DGs (directors general) as well as his own organisation”, commented one person. In a similar vein, another respondent argued that, “ADM (HR Mil) tries to pursue too many objectives”.

44th Annual Conference of the International Military Testing Association
Ottawa, Canada, 22-24 October 2002
7. Tied closely to the observed changes in leadership and leadership styles was a much complemented process of accountability. People now know who is responsible for what and who could be held accountable for actions and activities. “HR leadership today is not sweeping problems under their carpet – they want to know what the problems are and want to find solutions”. Indeed, one respondent felt that the consultative process around HR2020 is indicative of greater accountability and better leadership. This comment also highlights the view of many respondents that HR is no longer left to chance but is being treated seriously and in a strategic manner as the HR2020 process demonstrates.

*HR Strategy and Planning*

8. Many respondents felt that the lack of strategic HR planning was epitomised in the Force Reduction Program (FRP) which is seen as affecting the organisation negatively and for a lengthy period. “Due to the FRP, the CF lost many good people, the program was not rational and ultimately very counter-productive and had a very negative impact on the organisation”, commented one respondent. Today, however, most agree that, “There is a greater and more sophisticated understanding of market forces and trends”.

9. While, in general, respondents were complementary on the degree and focus of HR planning within the CF, some argued that the organisation is still slow to react to change and is still not doing a good job at foreseeing and pre-empting future threats. Furthermore, one person argued that, “The CF is not good at strategic planning. The organisation is always struggling with policy versus service delivery, strategic versus tactical, …so tempted to do tactical and not deal strategically”.

MANY OF THOSE INTERVIEWED OR INVOLVED IN FOCUS GROUPS BELIEVE THAT DND IS NOW BEGINNING TO MAKE ATTEMPTS TO IMPROVE STRATEGIC THINKING. ONE PERSON COMMENTED THAT “ONE KEY ORGANISATIONAL STRENGTH IS THE CURRENT FOCUS ON STRATEGIC PLANNING AND ALIGNMENT”. ANOTHER STATED THAT, “THE RECENT EVOLUTION OF THE HR PLAN … IS AN ATTEMPT TO BE FORWARD LOOKING”.

10. Almost all respondents agreed that HR is now being given the focus it deserves. Rather than being seen as an “add-on” as in the past, HR is recognised as a high priority and integral to all planning within the organisation. People also recognise that this refocusing takes time and there continue to be a number of issues and problems yet to be resolved.

11. During the interviews, one of the most immediate issues centred on the perceived lack of flexibility in HR management. One person felt that, “There is a liberal use of the word flexibility in this organisation to obtain privileges that really shouldn’t be given”. However, most other respondents argued that there is not enough flexibility. The common refrain was that “One size does not fit all”. Whether in terms of career paths, or educational and professional development, most of those interviewed believed the CF can better adjust to the needs of members while still achieving organisational goals.

*Culture*

CHANGES IN LEADERSHIP AND HR PRACTICES DO NOT TAKE PLACE IN A VACUUM. SUCH CHANGES TAKE PLACE WITHIN THE CONTEXT OF AN ORGANISATION WITH A STRONG CULTURE. FOR A NUMBER OF RESPONDENTS THE KEY TO ORGANISATIONAL CHANGE IS CULTURAL CHANGE AND OPINION WAS SOMEWHAT SPLIT ON WHETHER THIS HAS OCCURRED TO THE EXTENT NECESSARY.
12. On the one hand, some respondents felt that the relationship between subordinates and leaders has changed. The leaders themselves are now being trained in a different way with more grounding in values and modern managerial strategies and techniques. Some believed that as a result of this training and a more open organisation, the gap in values between Canadians and the CF has declined.

13. One way in which the culture of the CF has changed, some felt, is through greater diversity and the acceptance of women, aboriginals and visible minorities. “There is a better structure to support diversity, for example, the Diversity Council”, stated one person. However, many others argued that there is still a lot of traditional thinking and that the organisation has a long way to go to embrace real diversity.

14. On the negative side, many respondents saw continued problems related to the traditional military culture. A number of people pointed out that, “Women’s contributions are still discounted”. Still others felt that the CF is not as inclusive as it should be, still targeting white males over women and ethnic minority group members. “The CF is becoming increasingly non-representative of Canadian society even with legislation in place” commented one respondent.

15. The military culture, many respondents felt, continues to push its members beyond what is deemed reasonable. “There is a strong culture in the CF where … primacy of operations pervades every element”, stated one respondent in reference to the “Can Do” philosophy that has not changed even with increased Opstempo. These attitudes many believed reflect a lack of value and respect for members and their families.

People Focus

16. Many respondents praised the CF’s new focus on people. Examples of these positive changes towards members included increased interest in quality of life programs, pay raises, better health services and human rights initiatives. Respondents felt this contrasted with previous practices including the FRP, which treated the member as dispensable and ignored members’ families completely.

ONCE MORE, HOWEVER, RESPONDENTS FELT THAT MORE CAN BE DONE TO AMELIORATE THE SITUATION FOR PERSONNEL AND THEIR FAMILIES. IN PARTICULAR, MANY OF THOSE INTERVIEWED REFERRED TO THE LACK OF WORK/LIFE BALANCE AS INDICATIVE OF A CONTINUED DISREGARD FOR PERSONNEL. THE CF, SOME FELT, CAN TALK ABOUT PEOPLE COMING FIRST BUT HAS TO LEARN “TO PRACTICE WHAT IT PREACHES” OR SUFFER THE CONSEQUENCES. AS ONE PERSON PUT IT, “IF WORK/LIFE BALANCE IS NOT RIGHT, THE BRIGHTEST AND THE BEST WILL LEAVE”.

Communications

IN DEALING WITH PEOPLE IN AN HONEST AND OPEN MANNER, MANY RESPONDENTS ARGUED THAT THERE IS A NEED FOR BETTER COMMUNICATIONS. IN PARTICULAR, THERE APPEARS TO BE A PROBLEM IN OBTAINING INFORMATION. WHILE PROGRAMS MIGHT BE AVAILABLE THAT ADDRESS CONCERNS AND ISSUES, INFORMATION ON THESE PROGRAMS SOME FELT, IS NOT READILY ACCESSIBLE. “PEOPLE FEEL THEY HAVE DIFFICULTY GETTING THE RIGHT INFORMATION ABOUT SUCH THINGS AS COMPENSATION AND BENEFITS AND CAREER PROGRESSION”, COMMENTED ONE PERSON.

17. External communications were typically viewed by participants as poorly managed. In a number of interviews and focus groups, people commented on the fact that the CF is
not proactive enough in obtaining media coverage. Typically, only the negative is covered by the media when referring to the CF. For that reason, a number of respondents argued for a renegotiation of the relationship with the media.

**Compensation and Benefits**

18. In spite of the noted problems of access to information on programs, one issue that appears to have been well addressed by the CF, according to respondents, is that concerning wages and benefits. In particular, people believe the organisation has rectified the gap between the CF and the private sector that has served to lure people away from the Forces. “The CF has made significant inroads and created a level playing field between the public and private sectors”, stated one person. As another respondent saw it, “The CF seems to be focused on the real issues and trying to deal with them – comparability of compensation packages”.

19. For almost all those participating in the study, the days of poor wage levels are over and this is no longer an issue for members. However, in at least one focus group, participants noted that tying pay to rank is creating some problems. A number of NCMs who might prefer to remain in their current occupations have to seek rank promotions to get better salaries. Many would prefer to remain where they are as skilled trades people rather than becoming managers and supervisors and giving up their trade.

**Education, Career and Professional Development**

20. Many respondents see the CF as offering tremendous educational opportunities for members. For some this is the greatest strength of the organisation. For others, the emphasis on educational requirements has a downside. “Everyone is in the schoolhouse, and we wonder why there’s no-one on the ships”, commented one focus group participant, exemplifying this latter attitude. Others argued for a better balance between professional development and “… people being able to get their jobs done”.

21. Some respondents questioned the value and applicability of a formalized education to the military setting. Academic studies were not seen as necessarily providing the skills leaders need in the CF of today. To many, experience is being neglected in favour of academic studies. Given that many joined the Forces to accrue experience, it was argued that academic study may not be helping retention efforts.

22. Some respondents felt that the CF today does a better job at partnering with institutions of learning to help provide the training members need. This helps establish links with external institutions and means the CF does not have to provide certain training using its own internal resources. Still others felt that this is not being done enough and argued that comprehensive professional development strategies should include secondments to universities, colleges, industry and other government departments.

**ON THE QUESTION OF CAREER MANAGEMENT, MOST RESPONDENTS FELT THAT THEIR PERCEPTION OF WHAT A CAREER MANAGER SHOULD BE DOING DIVERGED FROM THE PRACTICE. RESPONDENTS EXPLAINED THAT CAREER MANAGERS WERE ALSO MENTORS, THERE TO HELP PROVIDE GUIDANCE AND ASSISTANCE. HOWEVER, IN PRACTICE, IT WOULD APPEAR THAT THESE MECHANISMS DO NOT WORK AND PEOPLE FEEL SOMewhat ABANDONED. FOR OFFICERS THIS IS LESS THE CASE THAN FOR NCMS. ONE RESPONDENT SUGGESTED THAT, “PROGRESS IN NCM PROFESSIONAL DEVELOPMENT HAS NOT BEEN GOOD AND IS NOT SEEN TO BE VERY STRONG, MOST LIKELY BECAUSE NCMS, AS A GROUP, ARE MORE DIVERSE THAN OFFICERS”.

44th Annual Conference of the International Military Testing Association
Ottawa, Canada, 22-24 October 2002
Recruitment and Retention

23. In almost all interviews and focus groups, respondents revealed an understanding of the relationship between all the various benefits, conditions of employment, programs, education and career development and the critical issues of recruitment and retention.

24. Many respondents pointed to the almost complete abandonment of recruiting during and for some time after the FRP. As one person put it, “The organisation has sucked the life out of recruiting due to them not having a long-term perspective”. According to another, “We lost control of the recruitment…. We cut all the people who did that work”. To rectify this situation it was suggested that a continued revamping of the HR management process and the military structure is required.

25. The CF needs to ensure that it remains competitive and attractive to new recruits. However, participants in one focus group adopted a somewhat traditional view and argued that closer attention needs to be paid to ensuring that the core values of the organisation are maintained and that recruits be selected on the basis of their adaptability to these values and to the military. As one person stated, “The CF is currently recruiting all the people who are suited to office work, exactly opposite to organisational needs…. The CF must recruit people who are, by their nature, adaptable to military society”. Even these participants, however, saw the need for programs and strategies to look after people once they are in so that they will remain in uniform.

THERE WAS A GENERAL UNDERSTANDING THAT MANY OF THE QUALITY OF LIFE AND OTHER INITIATIVES BENEFIT NOT ONLY RECRUITMENT BUT ALSO RETENTION. WITHOUT OFFERING PRECISE REMEDIES, MOST RESPONDENTS REFERRED TO THE NEED TO “DO THINGS RIGHT” AND NOT SIMPLY THROW MONEY AT THE RETENTION ISSUE IN ISOLATION.

26. Making the CF more attractive for newcomers and members alike and keeping pace with modern society includes allowing greater flexibility in terms of career management. Several respondents mentioned the notion of allowing for greater mobility in and out of the organisation during a member’s career. Some also pointed to the need for increased mobility between the regular and reserve forces as a way of keeping people interested and allowing for professional and personal development.

Reserves

27. On several occasions, respondents referred to the need to ensure that reservists are viewed as part of the Total Force. Some felt that distinctions have been made between regular personnel and reserves that diminishes the importance of the latter. Furthermore, some believed that reservists need access to the same services as other members of the CF.

28. Criticism was also leveled at the lack of reliable HR information on reserve members. A weakness, one respondent felt, is that the organisation has no common and comprehensive HR data system for regular and reserve forces. Time demands and expectations on reserves were also seen by some as unrealistic, given that they are part-time, and have other jobs and/or school.

THE CF IN THE FUTURE

General

29. Respondents and focus group participants were led through a discussion on the future of the CF and the role of HR in achieving that vision. Attention was focused on the actions and strategies people felt the CF should be engaged in to become the kind of force
required in the world of 2020. Respondents were also asked to think of possible “dark spots” or hidden problems that might get in the way of achieving these goals.

30. Using the same general sub-headings as in the previous section of the paper, the many and diverse comments have been synthesised and organized in a comprehensible fashion. From this analysis what becomes apparent is the uniformity of opinion in terms of where the CF should be headed and how it should get there. It also appears to be the case that most respondents feel the CF is generally doing the right things to achieve these long-term goals.

Leadership

31. The CF that many respondents see for 2020, will be a flatter, less hierarchical organisation where rank will be of less importance. As one put it, “Leadership and Officer/NCM distinction will fundamentally reshape the CF and possibly do away with Officer/NCM distinction”. Another respondent went further and explained what this change might involve, “Leaders will have to be comfortable operating in a much less rigid rank structure, where power will go up and down, and where we will have to be able to discuss and receive feedback”. Many others agreed that a new structure of leadership will inevitably mean greater openness and more two-way communications.

32. All respondents felt that the CF leaders of the future face a very different organisation than is in place today. They also believed that these future leaders will require different skills than the leadership of today. For most respondents, the CF leadership of 2020 will have developed a greater range of what one person called “softer skills”. These will include “…the ability to motivate, work in teams and be ethical”. This latter quality is one that was reiterated in several interviews and focus groups as a requirement. While some argued that ethics cannot be taught, most felt that greater attention to inculcating ethics into leadership skills is necessary to a future, more open CF, where leaders will be more “…accountable and responsible”.

33. Among the other qualities seen by many as not just desirable but essential in the leaders of the future is “thoughtfulness”. Some respondents felt the essential point here is the ability to think critically. The notion that the leaders of 2020, at all levels, will have to be “empowered decision-makers” was one subscribed to by many respondents.

34. One other important attribute of future leaders is the ability to communicate, not only internally, but also with external agencies and the wider population. The lack of what some termed “political astuteness” in dealing with the public and government, will have to be overcome through training of future leaders. Such training will also emphasize strategic and independent thinking.

HR Strategy and Planning

35. Many respondents argued that a more flexible organisation will be necessitated by increased competition between the CF and other public and private sector employers. “As an organisation providing HR functions, we will need to be extremely agile, with a solid strategic base/focus”, stated one person. HR strategy will also cover “the whole defence team which must include reservists”.

36. Some of those interviewed in the course of the study felt that to meet the needs of 2020, HR in the CF will have to become more professional with a greater, “…blend of military (for experience and understanding) and HR expertise (which could be civilian)”. A more sophisticated HR organisation will be “… a highly responsive, complex system
with the ability to develop our doctrine and concepts to be flexible to change”, contended one respondent.

37. Personnel policy must reflect the changing values and needs of individuals and cannot be “one size fits all”. Given that most respondents felt the CF will face increased competition in recruiting skilled people, policies need to be in place to allow for the optimisation of benefit packages and future employability.

Culture

38. Respondents felt that the HR changes, seen as necessary to face the world of 2020, will lead to a changed CF culture. A flatter and less hierarchical organisation meant for many respondents that there will have to be a change in mindset, values, and valuation of members, therefore calling for a change in organisational culture.

39. Flatter organisations require leaders who do not simply give commands. Leadership will be defined in terms of two-way communications and consultation within teams. Keeping pace with a changed society, leaders of the CF in 2020 will be “creative”, “at ease with cultural diversity” and respectful of individual member’s rights.

40. Developing this new culture will require a different kind of training, more focussed and consistent with “an overarching education vision”. Part of this change will also involve narrowing the perceived gap in values between the CF and the rest of society. It will be a role of future leaders to ensure that the organisation has credibility within the larger society by reflecting values relating to respect and “the freedom to question authority”.

People Focus

41. An essential part of the culture change perceived as necessary by many respondents, is a greater focus on people. There seemed to be an acceptance, by most in the study, that the target recruiting population of the future will be more individualistic and more rights oriented. Many of these people will also be in the enviable position of selling their skills in a seller’s market. For practical, as well as moral reasons, therefore, the CF will have to put people first, while ensuring the organisational goals and mission are met.

42. This new focus will be reflected in all aspects of the organisation and will require cultural as well as institutional change. Achieving these people goals will be the job of expert professional HR managers and flexible HR strategies.

Communications

43. Many respondents pointed to the need for greater consultation and better communications within the CF of the future. The changed structure of the organisation and the changed expectations of members will require that in some instances orders be replaced by consultation. This must be accompanied by transparency such that when decisions are taken, people know the “how” and “why” and feel that they were part of the process.

44. The future CF will be operating more closely with other public agencies and will be more directly competing for new recruits with the private sector. For these reasons, the organisation will have to acquire the ability to communicate directly with the public and with the media. This also means that the CF will be under greater public scrutiny and so must adhere to policies that enhance transparency and openness.

Compensation and Benefits

45. In many of the other areas of questioning, respondents referred to changes in the delivery and extent of compensation and benefits in the future. Whether discussing
leadership, HR strategies or recruitment, a comprehensive and integrated package of benefits is seen as essential to get the right people in the right position in the military of the future.

46. Respondents saw compensation and benefits as competitive with other employers and managed more directly by members. Personalised packages are viewed as a necessity allowing for greater flexibility. Benefits will be more inclusive, as they are extended to a greater degree to members’ families.

47. Pay will not be as directly linked to rank as in the past according to many respondents. Preferable for these people is a system that is linked to skill. This will allow members greater occupational flexibility with more lateral progression. Finally, these packages will also be available to those serving in the reserves as the distinction blurs between regular and reserve forces within the context of a Total Force.

*Education, Career and Professional Development*

48. For many respondents, one of the best selling features of the CF in 2020 will be that it is a learning institution. This means that people can be recruited who want to enhance or acquire new skills, outside of the conventional educational system. Indeed, one respondent felt that by 2020, such an education will be too costly for many and that the CF will offer a viable and attractive alternative.

49. Once in, members will receive expert career management. Although this technically happens today, some respondents felt that, “They are not career managed but rather vacancy managed”. As a way of retaining and recruiting people, a comprehensive career management process will allow people to see and realise a rewarding future in the CF.

50. In terms of professional development, many of those interviewed felt that there needs to be radical changes if the CF is to face the challenges of 2020. The selection and evaluation of people for promotion will be one area of change. As one person put it, reflecting the thoughts of others, “Currently the promotion board is a council of superiors, which leads to choices that mirror themselves. We will have to develop selection protocols pertaining to how we select and assess. Competition for military positions is an option”.

*Recruitment and Retention*

51. The CF of 2020 will require fewer but more skilled people who will expect flexible careers and a flexible organisation, according to many of those interviewed. Periods of engagement will be more flexible because as one person noted, “Youth of the future will not want long-term careers. Today periods of engagement are not flexible. We apply a template, which is no longer appropriate”.

52. As many people pointed out, we are recruiting today the leaders of the future. For this reason, “Bright, multilingual individuals are more important and more valuable. The idea of “cannon fodder” will totally disappear”. Furthermore, the increased diversity of society in 2020 will necessitate targeting an increasingly diverse recruitment pool today to enable the CF to better reflect the changing society.

53. Common to almost all responses on this question of recruitment in the future was the notion of a more sophisticated recruiting strategy. This strategy will endeavour to put the CF ahead of the curve. An important element of this strategy will be improved quality of life and personalised compensation and benefits packages.
DARK SPOTS

54. While organisations can plan for many eventualities, it is impossible to be prepared for every possible outcome. In seeking to achieve its long-term goals, the CF will be faced with some significant challenges that may prove difficult to overcome. In both the interviews and focus groups, people tended to identify a small number of very similar “dark spots” that they felt could derail HR planning for the future within the CF.

55. What Canadians think about the CF impacts on the political decision-makers. Given that public opinion can be fickle, some respondents felt that if public support wanes then so too will political support, putting funding in jeopardy. Allied with what some see as increased costs in maintaining competitive compensation and benefits, “the Forces may simply not be sustainable in the future.”

56. One of the most widely cited dark spots was related to the external world. The actual business of the CF will be determined not just by the leadership of the organisation but by public opinion and politics. Until we know what our business is going to be then we cannot properly resource to meet the mission. This makes effective HR management difficult if not impossible.

57. Another internal dark spot is related to the commitment of leadership to follow through with needed HR changes. Given that the CF leadership is viewed by many respondents as risk adverse, and that the leaders operate within an organisation that is “…resistant to change”, hoped-for change may not occur. Furthermore, more than one respondent talked about the lack of leadership commitment towards HR. Several people spoke of this perception echoed in the following quotations. “Currently leadership has a lack of commitment in advancing the HR agenda, which will lead to a lack of credibility for the CF”, stated one person. Another asserted that “In the CF, HR develops policy, plans, strategy, but doesn’t follow through and implement them.”

58. Given the importance of a change in culture as a necessary condition for the new CF of 2020, one frequently cited dark spot is the pervasiveness of “non-inclusive” attitudes within the organisation. The persistence of racist or sexist attitudes at all levels of the organisation will inhibit the growth of an organisation reflective of the wider society. This was tied in with the notion of reflecting wider societal values. “Reflecting Canadian values is both a mission imperative and is existence imperative…”.

CONCLUSION

59. The need for change was a theme that ran through all interviews and focus groups. This change must be more than superficial. Rather, almost all those participating in this study were of the opinion that real change is required to ensure the survival of the CF into the future.

60. Change is typically seen as being spearheaded by a strong leadership. This leadership must reflect cultural and organisational change through greater consultation and transparency. The vehicles for change must include recruiting strategies that reflect current and projected demographics of the country and new skill sets required for 2020 and beyond. Within the organisation, greater attention must be paid to building flexible and tailored career strategies and benefits packages.

61. A CF that is more diverse in terms of membership and agile in the way it operates will be the hallmark of an organisation prepared to face the challenges of 2020. Many people expressed misgivings about the ability of the organisation to rise to these
challenges. However, almost all felt that steps have been taken in the right direction and they are waiting to see if the leadership will continue on this path forward.

70. The consultations reported in this Research Note were part of a larger consultative process that also included the HR2020 Working Group and a number of presentations and information sessions. The feedback received was used to refine the HR2020 document. This document is in the process of being distributed throughout the CF. The next step in the continuing process of developing strategic HR policies and plans, is the identification of lower level strategies which will help us achieve the longer-term goals specified in HR2020. This work will include ongoing monitoring of external and internal driving factors, further consultation, and continued assessment and evaluation of strategies.
DIMENSIONS OF MILITARY LEADERSHIP RELATED TO SOLDIER WELL-BEING AND READINESS

Leora N. Rosen, Ph.D.
Senior Social Science Analyst,
National Institute of Justice,
810 7th Street N.W.,
Washington, D.C. 20531
rosenl@ojp.usdoj.gov

Kathryn H. Knudson, Ph.D.
Uniformed Services University of the Health Sciences,
Bethesda, Maryland

Peggy Fancher, M.SW.
U.S. Army Medical Command,
Health Policy and Services Directorate,
Ft. Sam Houston, TX

Note: The views are those of the authors and do not purport to reflect the position of the Department of the Army, the Department of Defense, the National Institute of Justice, or the Department of Justice.

ABSTRACT

Purpose:
The goal of this study was to describe and validate two new scales measuring dimensions of military leadership believed to have important implications for readiness and psychological well-being of servicemembers. The first scale, Motivational Leadership measures the degree to which small unit leaders are perceived as encouraging a positive work ethic among soldiers, emphasizing discipline, accountability, and respect. The second scale, Exploitive Leadership measures the degree to which unit leaders are perceived as self-serving, dictatorial, and unconcerned about the well-being of their subordinates.

Subjects and Data Collection:
Subjects comprised 1,190 active duty men and 206 active duty women who were stationed at an Army post in Alaska (referred to as Fort A), and who participated in a survey on family wellness in the summer of 1998. Battalion and family life chaplains administered questionnaires to soldiers in twenty-seven separate companies. The questionnaire included measures of psychological well-being, and accepted measures of military leadership and combat readiness -- the Walter Reed Army Institute of Research Vertical Cohesion and Combat Readiness scales. The questionnaire also included items pertaining to the two new hypothesized dimensions of leadership.

Analyses and Results:
A factor analysis confirmed that Motivational and Exploitive Leadership are separate factors, distinct from each other and from Vertical Cohesion. Motivational Leadership was positively correlated with Vertical Cohesion, and negatively correlated with Exploitive Leadership. Multivariate analysis revealed that Motivational Leadership was the strongest predictor of increased Combat Readiness. Exploitive Leadership was associated with decreased psychological well-being, but was not associated with Combat Readiness.

Military leadership plays a central role in building cohesion and combat readiness1, in promoting and maintaining the mental well-being of the soldier (Bliese & Halverson, 1996), in reducing stress (Dahl & O'Neal, 1993), and in preventing breakdown under stress (Solomon, Margalit, Waysman & Bleich, 1991). Supportive leadership has been shown to contribute significantly to soldiers' psychological well-being (Bliese & Halverson, 1996; Griffith, 1997), to family members' psychological well-being and satisfaction with military life6, and to the reduced likelihood of stress reactions following exposure to combat (Solomon et al., 1991; Solomon, Mikulincer & Hobfill, 1986).

Anecdotal information from military sources suggests that specific leadership qualities may prevent burnout among soldiers who face stressful conditions (Marlowe, 1986). These qualities include a reliance on competence rather than rank for achieving goals. Conversely, authoritarian leadership that uses rules and pressure to keep workers under control is associated with greater likelihood of burnout. Leaders who rigidly control the work environment and do not seem to care about their subordinates are believed to create an atmosphere of poor morale and disappointing productivity (Wilcox, 1994).

Consideration and structure are two elements of leadership that have been examined in relation to worker satisfaction. Leaders who are high in consideration emphasize the well-being of group members and create an atmosphere of trust, respect, and two-way communication. Those who are high in structure emphasize organized group activities to achieve goals. A study of nurses assessed the leadership style of the head nurse in relation to burnout among staff nurses (Duxbury, Armstrong, Drew & Henly, 1984). Low consideration among head nurses was associated with high levels of burnout. Structure itself was not related to burnout, but interacted with low consideration to produce high levels of burnout. Thus, head nurses who were high in structure but low in consideration had extremely high rates of burnout among their staff nurses.

Characteristics of supervisors may also ameliorate the effects of stressful work conditions, as suggested by a study of military nurses in which workplace stress was less likely to lead to burnout if the supervisor was supportive (Constable & Russell, 1986).

Much of military psychiatric research focuses on supportive leadership, or vertical cohesion, which is characterized by communication and social interaction with subordinates, knowledge of personal information about subordinates, listening to their opinions and caring about their families (Manning, 1994). Many studies have addressed the positive impact of supportive leadership, but few military studies have addressed the negative impact of authoritarian leadership beyond anecdotal description. The present study attempts to address this gap by developing new measures of leadership for use in a military context. The first deals with non-caring authoritarian leadership, which we call "exploitive" leadership. The second deals with leadership through an internalized moral
structure transmitted through example, which we call motivational leadership. These concepts were originally developed in order to study the relationship between leadership qualities and sexual harassment in the Army. The study found that exploitive leadership (referred to as self-serving leadership) was associated with increased perceptions of sexual harassment (The Secretary of the Army's Senior Review Panel, 1997). In a previous study, vertical cohesion had been associated with decreased perceptions of sexual harassment in the Army (Rosen & Martin, 1997). The present study examines two outcome variables in relation to three dimensions of leadership. The outcome variables include psychological symptoms of depression—a measure of mental health, and combat readiness, a measure of work functioning. The three dimensions of leadership include exploitive leadership, motivational leadership, and vertical cohesion. This study also controls for the effects of demographic variables and type of unit (combat versus combat support), and the interaction effects of type of unit by type of leadership on outcomes.

THE "FORT A" SURVEY

The sample for this study comprised 1,190 active duty men and 206 active duty women who were stationed at an Army post in Alaska (referred to as Fort A), and who participated in a survey on family wellness in the summer of 1998. Battalion and family life chaplains administered the questionnaires to soldiers in twenty-seven separate companies. Soldiers were informed that participation was voluntary and anonymous and were not asked for any identifying information either about themselves or their units. However, questionnaires from the same unit were kept together for the purpose of group level analysis. The focus of the study was on enlisted men and women. Officers were told that their participation was not essential because we wished to alleviate any concern that they might be recognizable by their demographic profiles.

Participants comprised about 58% of all non-deployed military personnel assigned to Fort A during the study period. On average, about 60% of the personnel from available units were provided with a questionnaire and an opportunity to complete it. Only about 3% of the questionnaires were returned blank, indicating refusals. Seventy percent of the men and 58% of the women were junior enlisted. Not unexpectedly, officers were under represented in the sample compared both to Fort A and the Army (Defense Manpower Data Center, 1998). Only 6% of the men and 2% of the women in the sample were officers compared to the population average of about 13% and 12% respectively. However, the rank profiles of the enlisted men and women in the sample were very close to those of the Fort A population. The study sample was also on average one year younger than the Fort A population (26 years versus 27 years), probably reflecting the under representation of senior-ranking officers.

The study sample and the Fort A population were highly comparable with regard to marital status profiles. In both groups 66% of men and 58% of women were married. Regarding ethnicity, there was a higher percentage of white men at Fort A compared to the Army as a whole (71% versus 62%). The study sample had the same percentage of white men as the Army. The proportion of women in the study sample and the Fort A population was similar to that for the Army as a whole (14.8%). There were also more
minority women in the study sample (36% white) compared to the Army average (43% white).

MEASUREMENT OF OUTCOME VARIABLES

The CESD Scale served as our measure of depression. This is a 20 item scale which has been widely used in epidemiological research on depression (Radloff, 1977). It has been shown to distinguish significantly between psychiatric inpatients and the general population, and has demonstrated good internal reliability. Its alpha coefficient in the present study was .91.

Combat Readiness is a measure of soldiers' confidence in their unit's training, readiness to perform its mission, and their fellow soldiers' ability to fight if and when necessary. It was one of several measures of unit climate developed at the Walter Reed Army Institute of Research (WRAIR) for the Unit Manning System Field Evaluation Study (Griffith, 1997; Marlowe, 1986; Griffith, 1988; Vaitkus & Griffith, 1990). It contains 8 items and in this study had an alpha coefficient of .84.

MEASUREMENT OF LEADERSHIP VARIABLES

The WRAIR Vertical Cohesion Scale is a measure of concerned leadership that was also developed during the Unit Manning System study, and has been used subsequently in studies of operational stress across deployments. It contains 9 items addressing the bonding between soldiers and their leaders at the small unit level, and soldiers' confidence in their chain of command. Its alpha coefficient in the present study was .87.

Motivational Leadership is a scale measuring the degree to which leaders encourage a positive work ethic among soldiers, by emphasizing discipline, accountability, and respect, by rewarding initiative and leading by example. It contains 14 items with an alpha coefficient of .91.

Exploitive Leadership is a scale measuring the degree to which unit leaders are self-serving and exploitative (The Secretary of the Army's Senior Review Panel, 1997). It contains 8 items with an alpha coefficient of .89. Items for the two new leadership scales, displayed in Appendix A, were factor analyzed together with items for the WRAIR vertical cohesion scale. The three factor solution, available from the authors upon request, confirmed that motivational and exploitive leadership are separate factors, distinct from each other and from vertical cohesion.

Type of Unit was dummy coded as 1=Combat; 0=Combat Support.

Demographic Characteristics included age, rank and gender. Gender was dummy coded as 1=Male; 0=Female.

ANALYSES AND RESULTS

The goal of this study was to examine the effects of three types of leadership on two outcome variables (depression and combat readiness) controlling for demographic variables, type of unit and the interaction of type of leadership by type of unit. Since the combat units in the study were entirely male, and since all the women in the study were
in the combat support units, the potential existed for the confounding of unit and gender. Therefore a new variable was created that incorporated unit type and gender. The sample was divided into three categories: (1) Women (n=206); (2) men in support units (n=722); (3) men in combat units (n=468). Differences in the study variables across these three categories are presented in Table 1.

Men in combat units were significantly younger than the women, who were significantly younger than men in support units. Men in support units were higher ranking and less depressed than the other two groups. With regard to combat readiness, motivational leadership and vertical cohesion, men in combat units had the highest scores, women had the lowest scores, and men in support units were in the middle. Women had higher scores on exploitive leadership than men.

Zero order correlations among study variables for the three groups are presented in Table 2. Motivational leadership and vertical cohesion were very strongly positively correlated with each other, and both were almost equally negatively correlated with exploitive leadership. Similarly, combat readiness was positively correlated with the positive leadership scales, and negatively correlated with exploitive leadership, while depression showed the opposite pattern. All correlations were significant except for that between age and combat readiness for women and for men in combat units. Combat readiness was positively correlated with age for men in support units.

The three leadership scales were examined as predictors of combat readiness and depressive symptoms using a single multivariate analysis of covariance model for multiple dependent and independent variables. Results are presented in Table 3. Only two variables emerged as significant predictors of combat readiness, namely, motivational leadership (f(13,1213)=60.4, p<.0001), and vertical cohesion (f(13,1213)=33.7, p<.0001). The strongest predictor of depression was lower rank (f(13,1213)=30.4, p<.0001), followed by exploitive leadership (f(13,1213)=35.6, p<.0001), lower vertical cohesion (f(13,1213)=10.1, p<.01), the interaction of gender/unit type by exploitive leadership (f(13,1213)=4.5, p<.05) and gender/unit type (f(13,1213)=3.4, p<.05). Post hoc comparisons of the three groups shown in Table 1 indicate that men in support units were less depressed than women and men in combat units. Zero order correlations in Table 2 indicate that exploitive leadership had a more negative impact on depressive symptoms among women than among men in combat units and in support units.

DISCUSSION

This study describes two new dimensions of military leadership that could have important implications for soldiers' psychological well-being and for readiness. Motivational leadership achieves its objectives through rewarding achievement, self-discipline, a strong work ethic, and moral accountability, including respect for others. Exploitive leadership achieves its objectives by instilling fear of punishment if the stated goals are not reached, with little regard for the methods used to reach them. Even though it may seem self-evident that the former style of leadership is desirable and the latter is undesirable, hitherto there has been little quantitative evidence to demonstrate the positive and negative effects of these dimensions of leadership on actual outcomes.
Furthermore, this study was able to show that each dimension of leadership has its major impact on different outcomes. Motivational leadership was a more significant predictor of combat readiness than vertical cohesion, yet both motivational leadership and vertical cohesion made distinct contributions to the explained variance in combat readiness, together accounting for 42% of the variance. Interestingly, exploitive leadership had no impact on combat readiness, yet it had a very large negative impact on soldiers' psychological well-being. Although vertical cohesion and motivational leadership had independent positive effects on well-being, the negative impact of exploitive leadership was far greater.

A significant interaction effect indicated that exploitive leadership had a more negative impact on the well-being of women than of men. Women may be more vulnerable to the negative effects of this style of leadership because of the added burden of gender discrimination. Leaders with exploitive styles may be particularly insensitive to issues of gender equity, and may thus have a more negative impact on women than on men.

In conclusion, the scales that have been described in this study may provide useful additional assessment tools in military research related to leadership, psychosocial stress, psychological well-being and readiness. These scales have been demonstrated to have good construct validity, excellent internal reliability, and the ability to discriminate between different important outcomes.

REFERENCES


Table 1: Study Variables Compared Across Type of Unit and Gender

<table>
<thead>
<tr>
<th>Variable</th>
<th>1 Women (N=206)</th>
<th>2 Men/Support (N=722)</th>
<th>3 Men/Combat (N=460)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waller-Duncan Mean(SD)</td>
<td>26.4 (5.8)</td>
<td>27.4 (5.9)</td>
<td>24.4 (4.1)</td>
</tr>
<tr>
<td>Agea</td>
<td>26.4 (5.8)</td>
<td>27.4 (5.9)</td>
<td>24.4 (4.1)</td>
</tr>
<tr>
<td>Paygradeb</td>
<td>4.3 (1.9)</td>
<td>5.1 (2.7)</td>
<td>4.5 (4.1)</td>
</tr>
<tr>
<td>Depressionc</td>
<td>0.9 (0.7)</td>
<td>0.7 (0.6)</td>
<td>0.8 (0.6)</td>
</tr>
<tr>
<td>Combat Readinessd</td>
<td>2.7 (0.7)</td>
<td>2.8 (0.8)</td>
<td>3.3 (0.8)</td>
</tr>
<tr>
<td>Mot. Leadershipe</td>
<td>2.8 (0.8)</td>
<td>3.1 (0.7)</td>
<td>3.4 (0.7)</td>
</tr>
<tr>
<td>Exp. Leadership</td>
<td>3.4 (0.8)</td>
<td>3.1 (0.9)</td>
<td>3.1 (0.8)</td>
</tr>
<tr>
<td>Vertical Cohesion</td>
<td>2.9 (0.8)</td>
<td>3.2 (0.7)</td>
<td>3.3 (0.8)</td>
</tr>
</tbody>
</table>

a) f(2,1384)=43.8, p=.000
b) \( f(2,1339) = 11.2, \ p = .000 \)
c) \( f(2,1331) = 10.9, \ p = .000 \)
d) \( f(2,1358) = 55, \ p = .000 \)
e) \( f(2,1375) = 41.9, \ p = .000 \)
f) \( f(2,1374) = 8.7, \ p = .000 \)
g) \( f(2,1358) = 18, \ p = .000 \)
Table 2: Zero Order Correlations Among Study Variables

<table>
<thead>
<tr>
<th></th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Rank</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>.50**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men/Combat</td>
<td>.44**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men/Support</td>
<td>.48**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Mot. leadership</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>.11**</td>
<td>.17*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men/Combat</td>
<td>.12**</td>
<td>.13**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men/Support</td>
<td>.18**</td>
<td>.27**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Exp. Leadership</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>-.17**</td>
<td>-.19**</td>
<td>-.80**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men/Combat</td>
<td>-.15**</td>
<td>-.18**</td>
<td>-.63**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men/Support</td>
<td>-.21**</td>
<td>-.33**</td>
<td>-.71**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Vertical Cohesion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>.12**</td>
<td>.13</td>
<td>.80**</td>
<td>-.69**</td>
<td></td>
</tr>
<tr>
<td>Men/Combat</td>
<td>.15**</td>
<td>.17**</td>
<td>.75**</td>
<td>-.56**</td>
<td></td>
</tr>
<tr>
<td>Men/Support</td>
<td>.18**</td>
<td>.25**</td>
<td>.77**</td>
<td>-.63**</td>
<td></td>
</tr>
<tr>
<td>6. Combat Readiness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>.03</td>
<td>.12</td>
<td>.62**</td>
<td>-.50**</td>
<td>.60**</td>
</tr>
<tr>
<td>Men/Combat</td>
<td>.06</td>
<td>.13**</td>
<td>.58**</td>
<td>-.38**</td>
<td>.56**</td>
</tr>
<tr>
<td>Men/Support</td>
<td>.12**</td>
<td>.16**</td>
<td>.59**</td>
<td>-.46**</td>
<td>.52**</td>
</tr>
<tr>
<td>7. Depression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>-.20**</td>
<td>-.23**</td>
<td>-.27**</td>
<td>.39**</td>
<td>-.24**</td>
</tr>
<tr>
<td>Men/Combat</td>
<td>-.21**</td>
<td>-.28**</td>
<td>-.22**</td>
<td>.26**</td>
<td>-.30**</td>
</tr>
<tr>
<td>Men/Support</td>
<td>-.19**</td>
<td>-.27**</td>
<td>-.23**</td>
<td>.31**</td>
<td>-.26**</td>
</tr>
</tbody>
</table>

** = p < .05
*** = p < .01
Table 3: Results of MANOVA Showing Effects of Dimensions of Leadership on Combat Readiness and Soldiers’ Psychological Well-Being

<table>
<thead>
<tr>
<th>Between Subjects Effects</th>
<th>Independent Variables</th>
<th>Multivariate Tests</th>
<th>Depression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Wilk’s Lamda</td>
<td>signif.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>signif.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>f value</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.7</td>
<td>p=.182</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p=.973</td>
<td></td>
</tr>
<tr>
<td>Rank</td>
<td>15.</td>
<td>p=.000</td>
<td>30.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p=.913</td>
<td></td>
</tr>
<tr>
<td>Gender/Unit Type</td>
<td>1.7</td>
<td>p=.135</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p=.892</td>
<td></td>
</tr>
<tr>
<td>Motivational Leadership</td>
<td>33.9</td>
<td>p=.000</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p=.000</td>
<td></td>
</tr>
<tr>
<td>Exploitive Leadership</td>
<td>17.8</td>
<td>p=.000</td>
<td>35.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p=.577</td>
<td></td>
</tr>
<tr>
<td>Vertical Cohesion</td>
<td>21.0</td>
<td>p=.000</td>
<td>10.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p=.000</td>
<td></td>
</tr>
<tr>
<td>G/U Type by Mot Lead</td>
<td>0.5</td>
<td>p=.733</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p=.413</td>
<td></td>
</tr>
<tr>
<td>G/U Type by Exp Lead</td>
<td>2.5</td>
<td>p=.04</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p=.560</td>
<td></td>
</tr>
<tr>
<td>G/U Type by Vertco</td>
<td>1.3</td>
<td>p=.256</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p=.396</td>
<td></td>
</tr>
</tbody>
</table>
Combat Readiness: $R^2=423$, $f(13,1213)=68.3$, $p<.0001$
Depression: $R^2=.164$, $f(13,1213)=18.2$, $p>.0001$
Appendix A

State your agreement with the following statements:
5=Strongly Agree; 4=Agree; 3=Not Sure; 2=Disagree; 1=Strongly Disagree

Motivational Leadership

1. The leaders in this company set high standards for all soldiers in terms of good behavior and discipline
2. The leaders in this company encourage soldiers to be all they can be.
3. The leaders in this company are able to take on tough problems
4. I am impressed with the quality of leadership in this company
5. My chain of command works well
6. The leaders in this company enforce the standards they set for good behavior
7. The leaders in this company take charge of things.
8. The leaders in this company set good examples for soldiers by behaving the way they expect soldiers to behave.
9. The leaders in this company are able to make tough decisions
10. The officers in this company would lead well in combat
11. The spouses of soldiers in this company would feel comfortable going to the leaders with a problem.
12. The leaders in this company wouldn't tolerate soldiers cheating on their spouses
13. The leaders in this company would take strong action against a soldiers who abused his or her spouse.
14. The leaders in this company keep soldiers well informed about Army programs available on this post to assist and support families

Exploitive Leadership

1. The leaders in this company are more interested in looking good than in being good.
2. The leaders in this company are self-centered.
3. The leaders in this company are bossy.
4. The leaders in this company push soldiers very hard to get things done without concern for the soldiers' needs.
5. The leaders in this company are not concerned with the way soldiers treat each other as long as the job gets done.
6. The leaders in this company just look out for themselves.
7. The leaders in this company are more interested in furthering their careers than in the well-being of their soldiers.
8. The leaders in this company are not concerned with soldiers' family problems.
MAXIMIZING THE SAILOR-RATING MATCH BY CONSIDERING APTITUDE AND INTEREST

Stephen E. Watson, Ph.D., David L. Alderton, Ph.D., & William L. Farmer, Ph.D.
Navy Personnel Research, Studies, and Technology Department (PERS-13)
Navy Personnel Command
5720 Integrity Drive, Millington, TN, USA 38055-1300
Bill.Farmer@persnet.navy.mil

Navy researchers, along with other contributors have recently developed new classification decision support software, the Rating Identification Engine (RIDE). The main goal behind the new system is to improve the recruit-rating assignment process so that it provides greater utility in the operational classification system. While RIDE employs many tactics to improve the assignment procedure, one additional strategy is to develop a measure of Navy-specific interests that may be used in conjunction with the current classification algorithm during accessioning. Job and Occupational Interest in the Navy (JOIN) is a computer-administered instrument intended to inform applicants about activities and work environments in the Navy, and measure the applicant’s interest in these activities and environments. It is purported that the simultaneous utilization of the JOIN interest measure and the RIDE ability components will improve the match between the Navy recruit’s abilities and interest, and ultimately serve as a means of increasing job satisfaction, performance, and retention.

The recruit typically has some degree of uncertainty when faced with the wide array of opportunities available from among more than 70 entry-level occupations, and over 200 program-rating combinations. The Navy, in deciding which rating is best suited for a recruit, must strike a careful balance between filling vacancies with the most qualified applicants and satisfying the applicants’ career preferences. Much is at stake in the process and research in civilian and military organizations has produced several pertinent findings. First, a lack of qualifications has been shown to lead to training failures and degraded job performance. Additionally, people who occupy jobs that are inconsistent with their interests are less likely to be satisfied with their work and are more prone to leave the organization for other job opportunities. Finally, dissatisfied
employees have higher absenteeism on the job, engage in more counterproductive behaviors, and seek alternative employment more often than their satisfied counterparts.

The JOIN vocational interest system will provide a critical component to the RIDE classification process. Current interest inputs to RIDE represent informal discussions with the classifier, which vary qualitatively by applicant. The JOIN system educates individuals about the variety of job related opportunities in the Navy, and creates a unique interest profile for the individual. The Sailor-Rating Interest fit for all Navy Ratings will be identified by comparing the Applicant’s Rating Interest Profile to each of the Rating Interest Profiles generated by JOIN. Once validated, JOIN will provide a standardized and quantified measure of applicant vocational interests, which will be provided as an input to RIDE. When proven successful, RIDE/JION can be implemented for initial classification, and transitioned to training and fleet commands for re-classification. Recent research efforts have focused on the development of the comprehensive JOIN Rating Interest Profile model for all Navy ratings, based on a series of analyses including iterative Subject Matter Expert (SME) interviews. Paralleling these efforts has been the development of the JOIN experimental software, also developed in concert with SMEs (see section below for details).

JOIN Model Development

The item development for the vocational interest inventory, Jobs and Occupational Interests in the Navy (JOIN), was an iterative process. The first challenge was to develop work activity and work environment items through an abbreviated job analytic procedure. A basic model of work served as the framework for the examination of Navy jobs and for the development of the inventory items. Conceptually, at the macro-level, the Navy consists of various job families or a grouping of jobs according to organizational function or work process (e.g., administration, health care, submarine personnel, etc.). Examining the world of Navy work on a micro-level reveals work activities or tasks that describe the work that is performed.
The first step in the item development process involved the collection of all of the available job descriptions from the Navy Enlisted Community Manager’s (ECM) website. A researcher reviewed each of these job descriptions, and highlighted words that reflected the following categories: 1) job families, or Navy community areas (i.e., aviation, construction, submarine, etc.); 2) work activity dimensions, or a process (verbs) and content (nouns); and 3) work context dimensions, or the work environment (i.e., working indoor, working with a team, etc.). From these highlighted extracts, lists of communities, processes, content words, and work environments, which seemed most representative of each Navy rating (N=79) were created. The process and content words were joined in various combinations to form process-content (PC) pair. These PC pairs would serve as individual interest items, allowing participants to indicate their level of interest in the work activity (e.g., maintain-mechanical equipment). Currently, a total of 26 PC pairs are included in JOIN.

After developing the content for the interest inventory, the next phase of the project was to design and create a computer-adapted measure of interests. The current version of the interest inventory, JOIN 1.01e, assesses three broad dimensions of work associated with Navy jobs and integrates over three hundred pictures of personnel performing job tasks to enhance the informational value of the tool. The first dimension, Navy community area, contains seven Navy community areas (e.g., aviation, surface, construction, etc.). Participants are asked to rank the individual areas, represented by four pictures each with its own text description, based on level of interest (i.e., most interesting to least interest, and not interested). The second dimension contains eight items describing work environments or work styles (e.g., work outdoor, work with a team, etc.). Participants are asked to indicate their level of preference for working in certain contextual conditions. Again, pictures with text descriptions represent each item. The final dimension, work activity, includes twenty-six PC pairs. Each PC pair serves as an individual interest item that allows participants to indicate their level of interest in the work activity dimension (e.g., maintain mechanical equipment, direct aircraft, etc.). Three pictures, and descriptive text, represent each PC pair item.
The computer-adapted version of JOIN is still in the developmental stages. Currently, a working version has been assessed for usability, clarity, and other characteristics that should promote ease of use and effectiveness as a measure of vocational interest in enlisted ratings (see section below for details).

JOIN Software Testing

Usability Testing I. The first test phase occurred during August of 2002 at the Recruit Training Center (RTC) Great Lakes, and was conducted with a sample of 300 new recruits (see section below for details). Participants were presented with JOIN and its displays, images, question content, and other general presentation features in order to determine general test performance, item reliability, clarity of instructions and intent, and appropriateness with a new recruit population for overt interpretability, required test time, and software performance. The initial results from the usability testing were very promising on several levels. First, the feedback from participants provided researchers with an overall positive evaluation of the quality of the computer adapted interest inventory. Second, the descriptive statistical analyses of the JOIN items indicated that there was adequate variance across individual responses. In other words, the participants were different in their level of interest in various items. Finally, the statistical reliability of the work activity items was assessed and the developed items appear to be very consistent in measuring participant interest in the individual enlisted rating job tasks. The results from this initial data collection effort have been used to improve the instrument before subsequent usability and validity testing.

Usability Testing II, Instrument Refinement. The second phase of testing is currently being scheduled. A group of 300-500 participants will be required to test the refined JOIN 1.0e instrument. These participants will allow data analyses to be conducted (e.g., usability, factor analysis, item reliability) on the refined JOIN instrument.
Criterion Related Validity Testing. Assuming satisfactory performance of the refined instrument, data collection will continue to establish criterion-related validity of the JOIN instrument. The predictive validation study will use new recruits who lack prior experience or knowledge of the Navy or Navy ratings. Criterion measures (e.g. A-school success) will be collected as participants progress through technical training, and those data become available. Participants’ social-security-number (SSN) will be collected in order to link interest measures to longitudinal data, including the multiple survey 1st Watch source data. Additional measures will also include attrition prior to End of Active Obligated Service (EAOS), measures of satisfaction (on the job and in the Navy), propensity to leave the Navy, or desire to re-enlist. Moderate and conservative estimates of sample size, based on power calculations of Chi-Square and Multiple Regression criterion validation methodologies, suggest a large sample of 3000-5000 is required to identify a predictor-criterion relationship if one exists. Roughly half of the sample will be used as the initial validation sample, with the remaining used as the cross-validation sample. It should be noted that the 300-500 participants used to assess the refined instrument would be used as part of this validation sample. Women and other minority groups will be over-sampled in order to provide sufficient sub-samples to perform analyses by sub-group. Participants will be asked to complete JOIN on laptop computers, which will take participants approximately 20-40 minutes to complete. All responses will be electronically collected for analysis.

In addition, following selected JOIN testing sessions, small discussion groups (with roughly 6-8 participants in each group) will be asked to provide feedback regarding their understanding of instructions, images, job descriptions, and area differentiation (i.e. “Community vs. Work-Style vs. Work Activity,” etc.). Discussion groups will be held immediately following test administration, with an estimated duration of 10-20 minutes. These groups will allow the content validity of JOIN to be assessed by the actual users of the system. Summary performance data and group discussion data will be analyzed and used to guide subsequent discussion groups.
NAVY PSYCHOMETRICS OF MEASURES – NAPOMS

William L. Farmer, Ph.D. & Ronald M. Bearden, M.S.
Navy Personnel Research, Studies, and Technology Department (PERS-13)
Navy Personnel Command
5720 Integrity Drive, Millington, TN, USA 38055-1300

Walter C. Borman, Ph.D. & Jerry W. Hedge, Ph.D.
Personnel Decisions Research Institutes
100 South Ashley Drive – Ste 375, Tampa, FL, USA 33602
Bill.Farmer@persnet.navy.mil

As we enter the 21st century, the military is in the middle of a major transition. Due to changes in mission and technology, the organizational and occupational characteristics that have come to define the present day military are being overhauled. As a result of this process it is imperative to develop a full understanding of the role that enlisted personnel play in the “new military.” This role includes interface with systems, equipment, and personnel that may appear quite a bit different than in the past. What individual requirements will these players need to accomplish their mission? How will this translate to personal readiness? How will performance be defined and measured?

In addition to individual requirements for successful performance, a number of other factors play important roles in the success of selection and classification in the military. The military uses results from large-scale aptitude testing as its primary basis for making selections into the service. Following this initial selection, testing results are further utilized in making classification decisions. The mechanism for making a personnel classification is a somewhat complicated process that involves using a combination of individual ability, obtained from the aforementioned testing program, the needs of the Navy (regarding jobs that need to be filled and gender and minority quota requirements), and individual interest. Historically, interest has received the least amount of weight.

Following classification into a job, an individual will go through basic training, then proceed to the training school pipeline that has been proscribed for the particular assigned career path. After finishing the initial training pipeline, an individual will be put
on the job, complete the first term of enlistment, then reenlist or not. A number of other factors, in addition to the things an individual brings into the service, play a crucial role in how that individual perceives their military career and whether the choice to reenlist or not is made. Organizational variables have typically received little or no attention in the military services when evaluating reasons for attrition or retention.

Historically, the preponderance of military predictive validation work has centered on measuring success in basic and technical training. Job performance in the first-term of enlistment has been included as a criterion measure sporadically. However, because finding and training a 21st Century sailor will be much more complex and costly than it is today, success on the job beyond the first term of enlistment in the Navy will be increasingly important. The prediction of such long-term behavior as reenlistment and promotion rates will require the use of new sets of predictor variables such as measures of personality, motivation, and interest. To effectively use the variables to predict long-term performance, it will be crucial to better understand the work context for the future Navy, including the environmental, social, and group structural characteristics.

Ultimately, combining the personal and organizational characteristics should lead to improved personnel selection models that go beyond the usual vocational and aptitude relations, encouraging a closer look at theories of person-organization (P-O) fit (see Borman, Hanson, and Hedge, 1997).

Advances in the last decade or so have shown that we can reliably measure personality, motivational, and interest facets of human behavior and that under certain conditions these can add substantially to our ability to predict attrition, retention, and school and job performance. The reemergence of personality and related volitional constructs as predictors is a positive sign, in that this trend should result in a more complete mapping of the KSAO requirements for jobs and organizations, beyond general cognitive ability. One particularly promising approach to measuring individual differences in the interpersonal and personality areas is the situational judgment test (SJT). These tests are based in the premise that there are important and often subtle differences between the behavior of effective and ineffective persons as they respond to
problems or dilemmas confronted in the course of carrying out their job responsibilities and that such differences are reflected in their responses to similar situations presented in written form.

Research has demonstrated that short-term, technical performance criteria, particularly overall school grades, are best predicted by general intelligence while longer term, more differentiated criteria such as non-technical job performance criteria, retention, and promotion rates are better predicted by other measures, including personality, interest, and motivation instruments. In order to select and retain the best possible applicants, it would seem critical to understand, develop, and evaluate multiple measures of short- and long-term performance, as well as other indicators of organizational effectiveness such as attrition/retention.

In general, then, when one considers what attributes are most relevant to perform effectively in any given job, there are many from which to choose. The type of person characteristic viewed as important to success in a job may vary from situation to situation. For example, for a job or set of jobs, one may be most interested in choosing persons that have high cognitive ability, and care much less about their personality or interest patterns. In other situations the reverse may be true. For optimal assignment, it is necessary to somehow link the attributes to how necessary they are for effective performance in specific jobs or job types, and as attempts are made to expand the predictor and criterion space, it will be important to extend one’s perspective to broader implementation issues that involve thinking about classification and person-organization (P-O) fit. As organizational flexibility in effectively utilizing employees increasingly becomes an issue (e.g., workers are more often moved from job to job in the organization), the P-O model may be more relevant in comparison with the traditional person-job match approach.

Current Research Program
The individual requirements for successful performance will require a thorough understanding of predictors of success and their relationship with key criteria. This effort will lead to the development of new measures of aptitude, personality, and other cognitive and non-cognitive instruments. Prior to this; however, it is necessary to develop a nomological net between the critical constructs that will define successful performance, and current selection instruments. The knowledge gained from this effort would provide a foundation for future developmental endeavors and contribute a much-needed component to the scientific literature.

With this said, the focus of efforts for the first year of the current research program (FY-2002), were a detailed illumination of the current state-of-the-science relevant to the changing nature of jobs in the Navy, and how this will impact selection and classification. A major component of this effort will be a thorough literature review of predictors and criterion and the relationships between them. This will include (but in no way is limited to): a.) extending work that was accomplished as part of the Army’s Project A, b.) review of current models of job performance, c) review of the literature on cognitive and non-cognitive predictor measures, d) investigation of promising areas (e.g. the role of situational judgment) for increasing predictive ability and objectifying measurement, e) the role of organizational and attitudinal variables, and f) person-organization and person-job match.

The deliverables for this phase of the program will include a knowledge base that will serve as the bedrock for the instrument development efforts that will characterize the forthcoming year (FY-2003). This knowledge, and the activities that led to its acquisition, will be archived via technical report and peer-reviewed publication.

Reference

NAVY QUALITY OF LIFE SURVEY
RESULTS FROM PAPER AND WEB VERSIONS

Gerry L. Wilcove, Ph.D.
Navy Personnel Research, Studies, and Technology
Navy Personnel Command, Millington, TN, 38055 USA
gerry.wilcove@persnet.navy.mil

BACKGROUND

The Navy Quality of Life (QOL) Survey has been administered periodically during the last eight years. “Life domains” (Campbell, Converse, & Rogers, 1976) such as Residence, Leisure and Recreation, Military Job, Relationship with Spouse or Intimate Other, and Standard of Living have formed the cornerstone of these surveys. In the previous Navy QOL surveys, everyone in the sample received a paper version through the mail (Craiger, Weiss, Butler, Goodman, & Wilcove, 1997; Wilcove, 1994; Wilcove & Schwerin, 2002). In the current survey, however, a Web option was also made available.

Although Web surveys have proliferated, research has not kept pace. Specifically, attention to potential differences between paper and Web surveys has been scarce. Miller, Kobayashi, Caldwell, Thurston, and Collett (2002) found demographic differences between respondents to paper and Web versions, but no differences in results or their variability. Lesser and Newton (2002) found higher response rates for paper than for e-mail versions, but no differences in item nonresponse percentages for closed-ended questions. Olmsted (2001) reported significantly more favorable responses on opinion scales from Navy Sailors completing the Web version than for Sailors completing the paper version. However, actual differences were exceedingly small in a practical sense.

PURPOSE

In this study, the paper and Web versions of the survey were compared on: (a) the proportion of Sailors completing the survey, (b), respondent demographics, (c) item nonresponse on potentially sensitive topics, and (d) results.

APPROACH

This year’s Navy QOL Survey addressed 13 life domains, including two new ones—Spiritual Well-Being (Vail & Gamerl, 2002) and Sailor Readiness. Independent samples were randomly drawn for 48 strata generated from four variables: paygrade band (e.g., E1 & E2, O1-O3), gender, minority/non-minority status, and Hispanic/non-Hispanic ethnicity. The Sample Planning Tool (Kavee & Mason, 2001) was used to determine how many Sailors to sample from each stratum.

A stringent criterion was established to determine whether or not individuals had completed the survey—the respondent must have answered one or more of the last five items. The demographic section had been placed last in the survey. Results were broken out into Sailors who had completed all of the five items, some of these items, or none of these items. A Cramer V statistic was computed for this 3 x 2 cross-tabulation matrix. The first two categories (completed all five items or some of these items) were then combined, yielding two categories (completed the survey, did not complete it) and a 2 x 2 matrix. This matrix was evaluated with the phi coefficient. Only those individuals completing the survey were included in the remaining analyses.
Demographic differences in the respondent samples (paper and Web) might lead to different results. For example, interviews aboard ship suggest that senior grade Sailors have easier access to the Internet than junior grade Sailors. Since meaningful, significant correlations are found between paygrade and opinions about QOL, paygrade and related variables were examined: enlisted vs. officers, time in the Navy (first term or obligation vs. subsequent duty), and paygrade band (e.g., E2 and E3). Similarly, opinions about QOL vary as a function of “family status.” Thus, an SPSS cross-tabulation analysis was conducted by mode across family status categories, such as married with and without children, divorced or legally separated with/without children in the household, and so forth. Given that separation from family and friends is a source of dissatisfaction, samples were also compared on the proportion of Sailors in shore billets vs. those in sea billets.

Depending on the demographic variable involved, a phi coefficient or a Cramer V statistic was computed to determine significance. Given significance, individual rows were examined to determine if percentages for the two mode groups differed by at least 10 points. To the researcher, 10 points—at a minimum—was a meaningful difference.

Relationship with Spouse/Intimate Other, Relationship with Children, and Standard of Living/Income were viewed as potentially sensitive topics. The Web version might, in turn, increase the sensitivity of such topics to individuals concerned about security issues. The researcher (and his colleagues) considered Standard of Living/Income as potentially the most sensitive topic, and thus this domain was placed after all other domains.

In the nonresponse item analysis, the percentage of nonresponses to the sensitive topics was compared by survey mode at the item level. It was noted that results could be confounded by the nonresponse of individuals who had already ceased completing the survey. The aforementioned “completion” criterion precluded that possibility, although more ideal approaches are available (e.g., including individuals completing items through Spouse/Intimate Other items who then quit filling out the rest of the survey). In the item nonresponse analyses, valid responses were recoded into one category (“responded to item”) and missing responses were coded into another category (“did not respond to item”). This resulted in a 2 x 2 (two mode levels) table that was evaluated with SPSS for significance using the phi coefficient.

The last issue was whether results differed by mode. A series of items (e.g., 8 items for Spouse/Intimate Other) addressed individual aspects of a domain (e.g., “How satisfied are you with the communication within the relationship?”). Aspects were grouped into scales or subscales on a conceptual basis, and Chronbach alpha internal consistency statistics were computed. Given acceptable results (? .70), a grand mean or (sub)scale score was computed, and a Student’s t analysis was conducted by mode. A computed effect size of ? .3 points between means operationally defined practical significance. If statistical and practical differences were found, ordinal regression analyses would be conducted to determine if differences were due, at least in part, to other variables besides mode—for example, paygrade.

Three items addressed overall QOL in the Navy. In addition, single items were used to ask Sailors how satisfied they were overall with their experiences in each of the domains. For both sets of items, the aforementioned procedures were employed to determine if results differed by mode.
RESULTS AND DISCUSSION
Completion of Survey
Table 1 presents results. A phi coefficient of .274, significant at the .01 level, was obtained. The greatest difference between the two groups lay in the percentage of Sailors who did not complete the survey—.5% of the paper version group vs. 13.7% of the Web version group. A total of 99.5% (54.7% + 44.8%) of the Sailors in the paper version group vs. 86.3% (47.1% + 39.2%) of the Web version group completed the survey under the criterion established for the study.

**Table 1**
Survey Completion

<table>
<thead>
<tr>
<th>Modes</th>
<th>Version</th>
<th>Web Version</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met criterion</td>
<td>54.7%</td>
<td>47.1%</td>
<td>52.4%</td>
</tr>
<tr>
<td>Minimally met criterion</td>
<td>44.8%</td>
<td>39.2%</td>
<td>43.1%</td>
</tr>
<tr>
<td>Did not meet criterion</td>
<td>.5%</td>
<td>13.7%</td>
<td>4.5%</td>
</tr>
</tbody>
</table>

Demographic Variables

No significant difference was found by mode in the number of enlisted relative to the number of officers. Approximately 57% of the paper version group and 56% of the Web version group were enlisted. Corresponding officer results were 43% and 44%. In contrast, the proportion of first-term/obligation personnel to other Sailors by mode was significant \( p < .01 \). However, the differences in response (yes/no) percentages were less than 7%. Having found, what the researcher considered to be, trivial differences between large groups of personnel (e.g., enlisted vs. officers), it seemed unlikely that within group differences would be found. Such was the case for paygrade bands. Although overall significance by mode was found \( p < .01 \), percentages differed at most by 3.8 points (9.6% of the paper version group were E1-E3 vs. 5.8% of the Web version group).

A significant \( p < .01 \) difference by mode was also found across family status categories. However, again, percentages differed at most by only 7.6 points (37.6% of the paper version group were married Sailors with children vs. 45.2% of the Web version group). Lastly, in spite of overall significance \( p < .01 \), a difference of only 8.9 points was found between the two mode groups in the percentage of personnel serving in sea and shore billets.

**Item Nonresponse**

As mentioned, Standard of Living/Income was placed last among the domains in constructing the survey because it was considered to be the most sensitive topic. A summary of item results for this domain is presented below. The results for Relationship with Spouse/Intimate Other and Relationship with Children were almost identical to results obtained for Standard of Living/Income.

The largest difference between mode item percentages for Standard of Living/Income was .8 points of a percentage (the percentage of Sailors not answering for one mode minus the corresponding percentage for the other mode). And, the largest number of missing responses, regardless of mode, was 1.4%. This result was found despite items asking Sailors how much they owed on personal unsecured and secured
debt, and whether they had experienced any financial crises (e.g., bankruptcy, repossesson of something purchased, or trouble paying child support payments).

Thinking the Sailors may have simply withheld information, two additional analyses were conducted. In these, Sailors were grouped into those revealing they had some amount of unpaid, secure or unsecured personal debt and those who stated they had no debt. For personal *unsecured* debt, 22.5% of the paper version group and 17.5% of the Web version group stated they owed money. The phi coefficient for the 2 x 2 matrix was significant (*p* < .01), but the difference in percentages (22.5% vs. 17.5%) was trivial. The same type of results were obtained for personal *secured* debt (26.4% for the paper version group vs. 21% for the Web version group, with an overall significant phi coefficient, *p* < .01).

Researchers have long been interested in the impact of asking for SSN on surveys, especially when addressing sensitive issues. The SSN issue resists a single hypothesis because of the complexity of human dynamics. For example, individuals may have no compunction about providing their SSNs, but may be leery of answering questions about their finances. Or, they may feel uncomfortable about providing their SSNs, but having left that item blank, may feel secure enough to answer sensitive questions. Additional hypotheses suggest themselves.

Since the current study found no important differences on sensitive topics by survey mode, it seemed unnecessary to conduct an ordinal regression analysis entering the SSN variable (provided, not provided) first. To accumulate basic information about Web surveys, however, it would be useful to know whether the SSN variable differentiated between the two modes. Results are shown in Table 2. Significantly more Web group respondents than paper group respondents provided SSNs (61.4% vs. 55.5%)(*p* < .01), but the difference was only 6 percentage points.

**QOL Perceptions**

In all, t-tests were conducted for 24 QOL (sub)scales, including the two measuring overall QOL and one measuring organizational commitment. Fourteen of the 24 were significant at the .01 level. However, none of the differences between mode means exceeded .24, less than the required effect size of .3. Additional t-tests were conducted on two outcome variables: long-term career intentions and plans to stay in the Navy or go at the next decision point. The t-value for the long term career intentions item was significant at the .01 level, but the difference in mode means did not approach the required effect size.

**Table 2**

<table>
<thead>
<tr>
<th>SSN Provided or Not Provide by Survey Mode</th>
<th>Mode</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Paper Version</td>
<td>Web Version</td>
</tr>
<tr>
<td>Provided SSN</td>
<td>55.5%</td>
<td>61.4%</td>
</tr>
<tr>
<td>Did not provide SSN</td>
<td>44.5%</td>
<td>38.6%</td>
</tr>
<tr>
<td><em>n</em> = 3,584</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>n</em> = 1,530</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>N</em> = 5,114</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusion

Provided that individuals are given a choice and freely choose to complete a survey on the Web, their data may differ little from Sailors who choose to complete the paper version.

References


Vail, T. C., & Gamerl, J. (2002). A pilot study to examine the use and relative validity of the Spiritual Identity Questionnaire in a deployed environment. Unpublished manuscript.


EQUAL OPPORTUNITY: ARE DIFFERENCES IN THE EYE OF THE BEHOLDER?

Stephen A. Truhon, Ph.D.
Department of Social Sciences
Winston-Salem State University
Winston-Salem, NC USA 27110
truhons@wssu.edu

ABSTRACT
It is an often-reported finding that minority group members report that equal opportunity climate as being less favorable than majority group members do. This finding has given rise to the question whether this difference reflects differences in the experiences of minority groups or whether there is a difference in how the groups perceive equal opportunity. The responses of 1,000 randomly selected American Indians/Alaskan Natives, Asians/Pacific Islanders, Blacks, Hispanics, and Whites to three scales (Differential Command Behavior toward Minorities and Women, Positive Equal Opportunity Behaviors, and Racist/Sexist Behaviors) from the Military Equal Opportunity Climate Survey were examined. Using item response theory (IRT), the item characteristic curves for each item by ethnic group were calculated. Results confirmed that Whites viewed equal opportunity in the military most favorably; American Indians/Alaskan Natives viewed equal opportunity least favorably. These results are further examined making use of differential item functioning and differential test functioning from IRT.

INTRODUCTION
A common finding in equal opportunity research is that minority group members find that equal opportunity (EO) climate as being less favorable than majority group members (e.g., Dansby & Landis, 1991). This finding has given rise to the question whether this difference reflects differences in the experiences of minority groups or whether there is a difference in how the groups perceive equal opportunity.

Several studies examining this question have employed confirmatory factor analysis (CFA) with conflicting results. McIntyre (2000) found that for three scales on the Military Equal Opportunity Climate Survey (MEOCS) there was a lack of construct equivalence. This would suggest that there is a difference in how the groups perceive EO. Bergman, Palmieri, Drasgow, & Ormerod (2001) examined scales of racial ethnic harassment and discrimination from the Equal Opportunity 1996 Status of the Armed Forces Survey (Armed Forces 1996 Equal Opportunity Survey, 1999) and found differences in harassment and discrimination but construct equivalence for the measures.

An alternative to confirmatory factor analysis in comparing groups’ responses to items has been derived from item response theory (IRT). IRT can examine differential item functioning (DIF; sometimes called “item bias”) and differential test functioning (DTF, sometimes called “test bias”). DIF provides information whether groups differ statistically (e.g., in their means) on an item; while DTF occurs when there are enough items with DIF to
produce a test that is not equally valid for the groups (see Shealy & Stout, 1993).
METHOD

Participants
The database for the MEOCS Standard (as of May 2002) was used for these analyses. All cases with missing data from the variables below were eliminated. There were 28,515 American Indians or Alaskan Natives, 45,376 Asians or Pacific Islanders, 171,198 Blacks (not Hispanic), 89,042 Hispanics, and 551,387 Whites (not Hispanic). From these cases, 1,000 cases were randomly selected from each ethnic group for further analysis.

Materials
Three scales were chosen from the Standard MEOCS: Differential Command Behavior toward Minorities and Women, Positive EO Behavior, and Racist/Sexist Behavior. The Differential Command Behavior toward Minorities and Women scale has nine items; the Positive EO Behavior scale has 11 items, and the Racist/Sexist Behavior scale has eight items.

Procedure
There were three steps in the major analyses. The first step involved the use of IRT. For these analyses, Thissen’s (1991) MULTILOG program was used to calculate the item parameters. Because the MEOCS uses a Likert-type rating scale, Samejima’s (1969, 1997) graded response model was employed.

In the second step, the parameters calculated on the same scales for each group must be equated with a common metric (using Baker’s [1992] EQUATE program). In the equating process, Whites were considered the reference group; all the other racial/ethnic groups were considered focal groups. The third step in the process involved calculating DIF using SIBTEST (Shealy & Stout, 1993).

RESULTS
Multivariate analyses of variance were performed on these three scales. As expected, there were significant differences between these racial/ethnic groups on the three scales (for Differential Command Behavior toward Minorities and Women, multivariate $F = 14.979 \, df = 36, \, 19960 \, p < .001$; all univariate $F$’s were significant at $p = .001$; for Positive EO Behavior scale, multivariate $F = 7.837 \, df = 44, \, 19952 \, p < .001$; and for Racist/Sexist Behavior, multivariate $F = 8.860 \, df = 32, \, 19964 \, p < .001$). In general Whites perceived fewer problems for these scales. For some items, Blacks and Hispanics report more problems in these areas than any other minority group.

IRT Analyses
Differential Command Behavior toward Minorities and Women
The item parameters (i.e., difficulty and discrimination) for these racial/ethnic groups were very similar, after they were equated. This similarity is illustrated in Figures 1 and 2 by the item characteristic curves for two items from this scale: MEOCS 10 (A majority supervisor frequently reprimanded a minority subordinate but rarely reprimanded a majority subordinate.) and MEOCS 18 (A majority supervisor did not select a qualified minority subordinate for promotion.).
Positive Equal Opportunity Behavior

The item parameters for these racial/ethnic groups were very similar, after they were equated. This similarity is illustrated in Figures 3 and 4 by the item characteristic curves for two items from this scale: MEOCS 7 (Majority and minority personnel were seen having lunch together.) and MEOCS 19 (When the Commander/CO held staff meetings, women and minorities, as well as majority men, were asked to contribute suggestions to solve problems.).

Racist/Sexist Behavior

The item parameters for these racial/ethnic groups were very similar, after they were equated. This similarity is illustrated in Figures 5 and 6 by the item characteristic curves for two items from this scale: MEOCS 6 (A majority first-level supervisor made demeaning comments about minority subordinates.) and MEOCS 12 (A group of majority and minority personnel made reference to an ethnic group other than their own using insulting ethnic names.).
Figure 2
Item Characteristic Curves for MEOCS 18 by Racial/Ethnic Group

Figure 3
Item Characteristic Curves for MEOCS 7 by Racial/Ethnic Group
Figure 4
ICCs for MEOCS 19 by Ethnic Group

Figure 5
ICCs for MEOCS 6 by Ethnic Group
Differential Item Functioning between Whites and Blacks

The difference between the item characteristic curves was greatest when comparing Blacks and Whites. Because SIBTEST (Shealy & Stout, 1993) can only compare two groups at a time, these two groups were compared for the following analyses.

Differential Command Behavior toward Minorities and Women

Differential item functioning analyses for the Differential Command Behavior toward Minorities and Women items were performed. Because nine comparisons are made, significance level should be set at .005 (.05/9). Despite the similarities mentioned above there are significant DIF for five, possibly six, of the nine items on this scale.

Positive Equal Opportunity Behavior

Differential item functioning for the Positive Equal Opportunity Behavior items were performed. Because 11 comparisons are made, significance level should be set at .005 (.05/11). Only two of the 11 items on this scale show significant DIF when comparing Whites and Blacks.

Racist/Sexist Behavior

Differential item functioning for the Racist/Sexist Behavior items were performed. Because eight comparisons are made, significance level should be set at .006 (.05/8). Five of the eight items show significant DIF when comparing Whites and Blacks.
DISCUSSION

The results of this study are somewhat in conflict. A visual examination of Figures 1 through 6 suggests that there is little or no DIF. Yet the results from SIBTEST indicate that there is DIF for almost half of the items. How does one account for this disparity?

One problem may with the linking factor to transform the parameters to a common metric. The EQUATE program (Baker, 1992) is used in parametric DIF analyses. Researchers using SIBTEST, a nonparametric DIF analysis, make use of a regression correction (Jiang & Stout, 1998). This could be a situation where a nonparametric test finds significance while a parametric test does not.

The author has not had a lot of experience with SIBTEST. Consequently, the results presented above should be tentative. It is advisable to compare the results obtained with other DIF programs (e.g., DFIT; Raju, van der Linden, & Fleer, 1995).

A third possibility is that the contradictory findings are both true. The items function very similarly for all racial/ethnic groups, but the difference in their functioning is enough to be statistically significant. In other words, the differences may be statistically significant but not practically significant. Further work on the MEOCS using IRT and DIF will help answer this question.

IRT and DIF provide tools for the further development of the MEOCS. This report focused on its usefulness in examining differences between racial/ethnic groups. However, the same approach could be used to examine differences between men and women with respect to sexual harassment and discrimination, and between senior leaders and enlisted personnel for several aspects of EO. IRT shows great promise in developing items to employ in an updated MEOCS. In addition, IRT is the underpinning for computerized adaptive testing, a possible technique for administration of the MEOCS.

References


AN INNOVATIVE WAY OF ADDRESSING
AGE-OLD PROBLEMS IN HUMAN RESOURCES

Luminita Stemate, Ivan Taylor
Director General Operational Research
National Defence Headquarters, Ottawa, Canada, K1A 0K2
luminita@ora.dnd.ca    iwtaylor@ora.dnd.ca

INTRODUCTION

The motivation for the work presented here was two-fold. First, the authors were confronted with the problem of finding ways to keep trained military personnel from quitting the forces in situations where their skills are very much in demand in the civilian world. Second, the authors were exposed to The Knowledge Wizard™ (KW) software package developed by Ideation International Inc. [1], and were interested in testing this software on a real problem. This paper focuses on the KW methodology by demonstrating its use on the retention problem in a generic military occupation. After a brief description of the software, the potential benefits and limitations of the software are discussed, followed by a case study and finally by a section dedicated to lessons learned and concluding remarks.

THE KNOWLEDGE WIZARD SOFTWARE

The KW software is an implementation of the “Theory of Inventive Problem Solving”. This theory emerged from a study performed over the course of 50 years in Russia, by Genrich Altshuller and his students [2,3]. They studied approximately 2 million patented inventions and innovations and observed that for every truly new and great idea, there are a multitude of “variations on the same theme”, all of which were granted patents. Based on this observation, they concluded that there must be a way to systematically enhance one’s creativity up to the point where obtaining a patent is made possible. In other words, while leaving the great inventions to chance discoveries by the geniuses of this world, there is still the possibility for the rest of us to produce “small inventions”, when properly assisted by a tool built for this purpose.

The KW is meant to be such a tool. This software package consists of a comprehensive database containing “best practices”, plus a set of algorithms whose role is to extract information from this database without the user having to perform queries. The set of algorithms that is added on top of the database also helps to make the system interactive to a certain degree. In a nutshell, the result could be described at the macro-level as a “user friendly database”.

At the micro-level, the KW process could be described through its various phases:
1. **Questionnaire.** The first phase of the process is a questionnaire that is made available by the KW software and which basically provides a structured framework that will allow the user to produce a good description of the problem he/she intends to solve. The questionnaire addresses issues such as the purpose of
the project, the problems expected, the resources available, past solutions to the same problem (if any), etc. Thorough completion of this questionnaire leads to a detailed problem definition, which is a necessary step towards the successful resolution of any problem.

2. **Diagram (Model).** The second step is to build a diagram (or model) of the problem. The detailed problem definition obtained after answering the questionnaire is reviewed with the purpose of clearly identifying the main goals and drivers of the problem. A diagram is then built, using some pre-defined symbols, to show the goals of the project together with the most important factors that have an impact (either positive or negative) on reaching these goals. The purpose of the diagram is to obtain a definition of the problem, simple enough to work with, while still covering all the main drivers of the problem and the relationships between them. Note that, although it might be tempting to skip Phase 1 and start the process with this simplified model of the problem, it is important not to do so, to avoid the danger of over-simplifying things. It is crucial to begin with a problem as well defined as possible, with as many details as possible, and then work your way towards the simplified model by “trimming” the original definition, cutting off the less important elements of it. Deciding what will go into the diagram and what will be left aside is a very important decision and following the process suggested by the KW can help make this decision defendable. It also increases the confidence that the model built is credible and valid.

3. **Directions for Innovation.** Once the model is built, the software will provide the user with a number of “directions for innovation”, which are directly based on the model. The software isolates elements of the diagram, based on the relationships between them, and creates directions for further study, or sub-problems on which the user should focus. The result of this phase is that the original problem is divided into a number of significant sub-problems. The underlying idea being the “divide and conquer” strategy. If the number of directions for innovation generated by the software is high, a prioritization of these directions is suggested. Although some guidance is offered regarding the rules for the prioritization, these rules are ultimately up to the user and they depend heavily on the time available for the project. If there is not enough time to cover them all, only a subset of directions (the ones showing more promise toward solving the problem) will be selected for further consideration. Each of the selected directions for innovation is to be tackled separately by the user, with assistance from the software.

4. **Examples from Other Fields.** The type of assistance that is offered by the software takes the shape of appropriate examples from other fields, provided to the user for each direction of innovation. The user is then supposed to use analogy to generate solution ideas for his/her own problem (or sub-problem, to be more precise), based on these examples. At the end of this phase, the user is in possession of sets of ideas that have potential to solve or improve one particular aspect of the original problem. These ideas are referred to as “preliminary ideas”.

5. **Combine Ideas into Concepts.** The purpose of this step is to make the set of preliminary ideas, which could be of a significant size, more manageable. Ideas are to be grouped in categories or concepts, based on the existent relationships
between the ideas. An additional goal of this step is to remove any redundancy in
the list of ideas.

6. **Evaluate Results.** The previous step is one way to make the ideas easier to
evaluate, by evaluating a whole set of ideas that are grouped together into a
concept. Concepts are to be evaluated with the purpose of selecting one or more
of them for implementation. As such, the evaluation process should cover issues
such as strengths and weaknesses of the concept, potential secondary problems
that might appear should the concept be implemented, etc. Costs, risks, benefits
and all other relevant factors should be considered as potential criteria to evaluate
the concepts. Note that, although the KW does provide some guidance for
evaluating the results, the responsibility of this step is largely placed upon the
user.

7. **Plan Implementation.** Similarly, the KW does provide some guidance with respect
to producing an implementation plan for those concepts that “passed” the
evaluation phase. However, this guidance is limited to a checklist of things that
any good implementation plan has to cover, which is useful, but, once again, the
responsibility for this stage of the problem solving process is placed on the user.

**ROLE OF THE KNOWLEDGE WIZARD IN THE PROBLEM SOLVING
PROCESS**

Figure 1 illustrates the authors’ view on how the KW fits into the problem
solving process. The various stages in this process were identified as follows:
problem definition, problem analysis, development of options, option analysis
(Benefits/Costs/Risks Analysis) and recommendations (including the
implementation plan).
It is the authors’ opinion, based on our experience with the KW, that although this software can be instrumental in the stages of problem definition, problem analysis and the development of options, it can only assist to a limited degree in the next stages of the problem solving process. Moreover, unless the problem to be solved is particularly simple or very qualitative in nature, the use of the KW should be complemented by the use of some type of quantitative tools, especially in the ‘options analysis’ stage. The black arrows in Fig. 1 are used to illustrate a potential major impact, while the grey arrows illustrate a limited impact that either the KW or what was generically called “Quantitative Tools” may have in the various stages of the problem resolution process.

CASE STUDY: THE RETENTION PROBLEM

The authors used the framework provided by the KW to address the problem of improving retention in a generic military occupation characterized by the fact that its members possess skills that are in significant demand in the civilian world.

During the phases of problem definition and problem analysis, the KW was the only tool utilized. The diagram (model) produced and used as the simplified version of the problem is shown in Figure 2.
In Fig. 2, the two objectives of producing and retaining the desired number of personnel are represented in the two boxes that do not have rounded corners. The boxes with rounded corners represent various other factors that were considered to have a significant direct or indirect negative impact on the two objectives. An arrow between two boxes indicates a positive cause and effect relationship between the two events. A crossed arrow signifies a negative impact.

After building the diagram in Fig. 2, the software generated 15 ‘Directions for Innovation’. Two of them did not make sense and were not explored. The 13 remaining directions were tackled one by one, and, based on the numerous examples from other fields that were provided for each direction, a team of two analysts generated 303 raw ideas over 16 two-hour sessions.

An example of one of the more creative ideas that were generated using the KW was the following. Based on the Direction for Innovation: Find an alternative way to produce the desired number of personnel that does not require too much recruiting and is not influenced by the inability to forecast requirements or too little recruiting. The KW software suggested: To ensure an organization’s growth, create a positive (reinforcing) feedback loop. In a positive feedback loop, every step in the organization’s evolution increases its value in the eyes of society and its members (or increases its market value), which stimulates further growth, and so on. The KW software provided the specific example: “Pyramid” schemes (illegal in many countries) are revenue systems that require an endless stream of recruits for success. Recruits (a) give money to recruiters and (b) enlist fresh recruits to create a bottom-up revenue source. The result is almost always inevitable: at best, the top-most members walk away with a lot of money, while most recruits lose their investments. The idea that we generated was: use the military training process to track the personnel that have gone through it according to who was
their mentor/instructor and compensate/reward these mentors/instructors later in their
career and maybe even after their retirement for the accomplishments of their students.

The 303 ideas were rated separately by the two analysts, using a scale of 1 to 4. The highest rating (‘4’) was awarded if the idea was considered a “Go” (i.e., the idea is definitely worth considering for implementation). The lowest rating (‘1’) was attributed to ideas that were considered “NoGo” (better to be abandoned). A rating of ‘2’ indicates a “Maybe NoGo” idea, while a rating of ‘3’ indicates a “Maybe Go” idea. The results of these ratings (illustrated in Table 1) were used to perform a first screening of the ideas. Only those ideas rated “Go” by both analysts or rated “Go” by one analyst and “Maybe Go” by the other (the figures in bold font in Table 1), were kept for further analysis, while the rest of the ideas were screened out.

### Table 1. First Screening of the Ideas

<table>
<thead>
<tr>
<th>NoGo</th>
<th>Maybe NoGo</th>
<th>Maybe Go</th>
<th>Go</th>
</tr>
</thead>
<tbody>
<tr>
<td>Go</td>
<td>1</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Maybe Go</td>
<td>8</td>
<td>16</td>
<td>30</td>
</tr>
<tr>
<td>Maybe NoGo</td>
<td>5</td>
<td>15</td>
<td>18</td>
</tr>
<tr>
<td>NoGo</td>
<td>5</td>
<td>12</td>
<td>5</td>
</tr>
</tbody>
</table>

The 148 ideas kept for further analysis were classified by their potential to lead to benefits in the short, medium or long term and then by the specific categories of solution they represent, in the spectrum of potential solutions to the retention problem. The result of this grouping is shown in Table 2.

A ‘concept’ would then be formed (in this case) by the ideas included in the intersection between one of the categories on the left in Table 2 and one of those at the top in the same table. For example, the eight ideas included in Quality of Life / Short Term Benefits (the two grey cells in Table 2) would form such a concept.

After grouping the ideas into concepts to, the KW suggests that the concepts should be evaluated using some appropriate criteria. However, the authors did not feel comfortable with the idea of evaluating the concepts, because they did not find it obvious that if a concept “passed” the evaluation tests, that necessarily meant that all ideas included in that concept are good ideas that should be part of the recommendation package. Instead, the authors felt that a “cocktail” or mix of ideas (not necessarily inter-related) might be preferred to the use of concepts (which, by definition, is composed of inter-related ideas).

### Table 2. Grouping Screened Ideas into Concepts

<table>
<thead>
<tr>
<th></th>
<th>Short Term</th>
<th>Medium Term</th>
<th>Long Term</th>
<th>Unrated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Go-Go</td>
<td>Go-Maybe</td>
<td>Go-Go</td>
<td>Go-Maybe</td>
</tr>
<tr>
<td>Quality of Life</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Training</td>
<td>2</td>
<td>10</td>
<td>0</td>
<td>9</td>
</tr>
</tbody>
</table>
An evaluation of each idea was completed taking into account the potential risks, benefits and estimated costs using a rough order of magnitude scale of “low”, “medium” and “high”. This process identified the ideas that showed the most potential and, as such, constitute good candidates for inclusion in the package of recommendations. The ideas were sorted according to maximum potential benefit and minimum potential risk subject to a certain estimated budget. Figure 3 shows the results of this analysis. For any given budget, the “cocktail” of ideas that maximizes the benefits with minimum risk can be determined.

LESSONS LEARNED AND CONCLUDING REMARKS

Some of the lessons learned after our experience in working with the KW include the following (in no particular order):
1. Give the process the time and resources it takes. In our study, we had two analysts working for 16 sessions of a maximum of two hours each.
2. Preferred size of a team should be two persons. Having two people “bounce” ideas off each other is an excellent method of creativity enhancement.
3. The preferred team would be one Analyst and one Subject Matter Expert (SME). The software may be user-friendly enough to have two SME’s work together. In our study, we had two analysts work together. This was probably not optimal.
4. Preferred work session duration should be around 2 hours. The creative process can be quite strenuous. It is often better to take a break rather than push one’s self to the limit of one’s energy.
5. To maximize the impact, get involved in the early stages of the project. Our work came into the process too late to have significant influence with this occupational class. However, our generic results may be useful for other trades.
6. We feel that the KW software could be used in a workshop format to enhance creativity in larger groups. For particularly difficult problems, several teams could tackle individual ‘directions for innovation’ separately and come together to compare results. Time permitting, one could allow for a certain amount of overlap between the directions to be tackled by each of the teams involved.
7. Include quantitative analysis with the KW. Prioritization and Benefits/Costs/Risks analysis can be highly beneficial to the recommendation process.

8. Capture together the ‘direction for innovation’ => ‘suggested approach’ => ‘example’ => ‘idea generated’. This will be useful when building a case to support the recommendations to see the specific part of the problem that is being addressed, how it is being addressed and an well-known example of how the methodology has been used elsewhere.

9. Go through the entire process (adapted for your own needs if necessary), complete with quantitative analysis, before making formal presentations to the decision makers.

To conclude this paper, there are two important questions that still need to be answered. The first one is: “Did we solve the retention problem in this trade using the KW?” The short answer would be “No”. The long answer would be “No, but we made significant progress”. We unfortunately came into the process too late to have a significant influence over the recommendations package. However, we feel that many of the ideas we generated could be easily transferred to other trades in the Canadian Forces. The second question that is anticipated from every reader is: “Would we use the KW software again?” Our answer is “Yes”. We found it very helpful in the problem definition stage of the work. Also, the creativity enhancement aspect of the software pushed us ‘outside the box’ and kept us working through the various directions and suggestions until we were convinced we had given the problem significant coverage.

REFERENCES

1. See the Ideation International Inc. website for more information on The Knowledge Wizard. www.ideationtriz.com

2. Altshuller, Genrich; The Innovative Algorithm; Technical Innovation Center; March 1999.

3. Terninko, John; Alla Zusman and Boris Zlotin: Systematic Innovation: An Introduction to TRIZ (Theory of Inventive Problem Solving); St. Lucie Press; Boca Raton, FL; 1998.
FOCUS GROUP INTERVIEWS:
AN INTEGRAL PART OF SURVEY RESEARCH

Dr. Alma G. Steinberg

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333-5600
steinberg@ari.army.mil

This paper addresses how integrating focus group interview methodology with survey research can enhance the reliability, validity, and utility of research findings. It describes key issues using this integrated approach and highlights some methodological considerations when conducting focus groups with the military.

Advantages of using both survey and focus group methodologies

Survey and focus group methodologies have both advantages and disadvantages (see Fowler, 1993; Krueger, 1988; Morgan, 1993; Stewart & Shamdasani, 1990; Ward, Bertrand, & Brown, 1991). However, since these two approaches complement one another, using both can result in maximizing the advantages and minimizing the disadvantages of each methodology (Morgan, 1996).

Survey methodology is well-suited for:

(a) use with very large samples
(b) administration in widely dispersed geographical locations
(c) stratified sampling
(d) ensuring that the same questions are asked in the same way to all
(e) minimal training of data collectors
(f) automated input of responses to a database
(g) quantitative analyses

Focus group methodology provides additional advantages over and above those of surveys. For example:

(a) Focus groups can contribute to the enhancement of survey instrument reliability and validity. Their input can be taken into account in the design of survey questions. Also, focus groups afford the opportunity to pretest (and post-test) question wording and response categories for meaning, clarity, relevancy, and completeness.

(b) Focus groups provide the opportunity for individuals to raise issues that are not currently on surveys. At the same time, participants in focus groups often indicate that they appreciate the opportunity to be “heard” — in a way that they do not feel not when they respond to surveys (Morgan, 1998).

(c) Focus groups provide context to interpret survey results and enable an in-depth understanding of the issues (Morgan, 1998). It is often not enough to know how many
people feel a certain way (i.e., from survey results); one needs to know why people responded as they did to the survey and why people feel as they do. Also, the illustrations/examples that focus groups provide can be used to give context to quantitative survey results and help research sponsors to better understand the soldier perspective.

(d) Focus groups contribute to finding solutions. They allow looking at an issue from multiple perspectives at the same time. This often helps to delineate the issues more clearly and to determine factors that need to be taken into account in figuring out solutions. Ultimately these lead to actual utilization of the findings.

Methodological considerations

Focus groups are interviews, conducted for research purposes, with small groups of participants, and are led by a well-trained moderator (e.g., trained in group dynamics, interview skills, and research methods). The moderator raises the set of discussion topics, using a pre-established discussion guide (Morgan, 1998; Stewart & Shamdasani, 1990; Ward, Bertrand, & Brown, 1991). Participant responses constitute the qualitative data that are to be recorded accurately and analyzed. Thus, focus group interviews differ from individual interviews, in that the discussion that takes place is an essential element. They also differ from other scenarios in which discussions take place (e.g., work groups, team building sessions, committee meetings, informal griping sessions, sensing sessions, and problem solving meetings led by leaders).

When conducting focus group interviews within the military, it is important to keep in mind the organizational culture. Care needs to be taken in order not to communicate to participants an expectation that there are certain “right” answers to questions. Also, participants need to be sure that confidentiality is maintained and that there will be no repercussions from the answers they provide. These considerations are especially important in the military because the focus groups are conducted within an organizational context (as opposed to, for example, focus groups that are conducted by a marketing firm which brings a variety of individuals together to discuss a certain product). They affect many aspects of planning and conducting focus groups, such as: the selection criteria for the participants, the amount and type of information provided beforehand about the topics to be covered in the groups, the privacy of the room in which the groups will meet, and the introduction the moderator presents to the group.

Thorough training in focus group methodology is essential for personnel involved in designing, conducting, analyzing and/or reporting the results. This cannot be emphasized enough. Even if one has conducted other kinds of research or has had considerable experience as a military leader, one still needs to be trained in focus group methodology in order not to introduce bias and invalidate the results (Reviere, Berkowitz, Carter, & Ferguson, 1996). Ideally, the training would include: interview protocol development, understanding and controlling group dynamics, interview skills, recording data accurately, research methodology, and qualitative data analysis. Also crucial is the requirement that the researchers understand military language, culture, and customs.
It is important to ensure that research sponsors understand that when they request focus groups be conducted, they are asking for qualitative and not quantitative data. Otherwise, they may expect outcomes in terms of numbers, when it is the group discussions themselves that provide the data – not head counting (Morgan, 1998). The integrated approach of using both surveys and focus groups helps to solve this dilemma by providing both the quantitative and qualitative data.

**Conclusions**

Surveys provide quantitative data from large geographically dispersed samples in which the same questions are asked in the same way to all. Focus groups provide qualitative data for an in-depth understanding of the issues and possible solutions. Focus groups enhance the reliability, validity, and utility of surveys.

When surveys and focus groups are conducted in conjunction with one another, using appropriate research methodology, they can be used to:

- (a) provide scientifically-sound, and timely information to decision-makers
- (b) monitor soldier issues
- (c) conduct program or policy assessment
- (d) determine the validity of anecdotal information or opinions
- (e) track trends on a wide number and variety of issues
- (f) identify emerging issues
- (g) assess the impact of unexpected events
- (h) provide scientific data (for speeches, conferences, publications, educational events)

**References**


OPTIMAL PLACEMENT OF SOLDIERS INTO MULTIPLE JOBS
Amos Spector, Ph.D., M.Sc.
Netanya Academic College, 16 Kibbutz Galuiot St, Netanya 42365, Israel
Spector@netanya.ac.il

Oded Gogol, M.Sc.
Israel Port Authority, 74 Begin Rd, Tel Aviv, Israel
Odgogol@netvision.net.il

The Problem
In a typical selection situation there is one job with k openings, and there are n candidates for this job (n>k). On the basis of some selection variables, such as tests, we do one of the following: (1) fill each of the k positions with the best n\(^k\) individuals (“best” in expected performance, as predicted by the selection variables), or (2) pick the best k, provided that their predicted performance is above some predetermined minimum, or (3) admit any candidate whose predicted performance is above a minimum (k is not limited in this case). Usually, decisions are based not on predicted performance, but on some “fitness score”, which is a weighted sum of the selection variables. In a simple example: if Jane’s fitness score is 84, and Gene’s is 72, and there is one opening, then Jane will be chosen to fill it.

The situation is different when there are more than one job to be filled. The simplest example would be a situation in which one wishes to place two candidates to two jobs. The immediate question would be how to use their fitness scores (for each of the two jobs) to decide which one of them should be assigned to the one job, and which - to the other. In some cases, such as the one presented in Table 1, the decision is simple.

Table 1: A case with simple solution

<table>
<thead>
<tr>
<th></th>
<th>Job A</th>
<th>Job B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jane</td>
<td>88</td>
<td>82</td>
</tr>
<tr>
<td>Gene</td>
<td>76</td>
<td>72</td>
</tr>
</tbody>
</table>

However, with a minor additional complexity to the situation, table 2 presents a harder decision to make. In this example we have three candidates and three jobs, and we need one person in each job. Looking at table 2, it seems attractive to place Jane in job A, and Gene in B. But this will force John into job C, which seems like a misfit. Should we place John into A and leave Gene in B, Jane will be forced into C. Again, an undesirable placement.

Table 2: A case with no easy solution

<table>
<thead>
<tr>
<th></th>
<th>Job A</th>
<th>Job B</th>
<th>Job C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jane</td>
<td>88</td>
<td>82</td>
<td>70</td>
</tr>
<tr>
<td>Gene</td>
<td>72</td>
<td>84</td>
<td>70</td>
</tr>
<tr>
<td>John</td>
<td>75</td>
<td>60</td>
<td>60</td>
</tr>
</tbody>
</table>
If the problem is extended to hundreds of candidates, and dozens of jobs, it is clear that its size is unmanageable by a human decision maker. Further complexity will rise if the situation involves additional demands and constraints such as: Gene will operate satisfactorily only in jobs C, F and G; the average fitness scores in job A must exceed that of job D, etc. Clearly, problems of this nature require some systematic approach for their solution.

We faced a problem of this type in a project that involved placing hundreds of soldiers, inducted simultaneously, into a few dozens of different jobs. Placements (or ‘assignments’), were to be based on a number of variables at the individual soldier level, as well as needs, rules and requirements of units (jobs) into which soldiers were to be placed. The present paper will describe the way we dealt with that problem.

Possible approaches

The problem of placing multiple candidates into multiple positions has been handled in the past (Cascio, 1998; Ghiselli, 1956). Three of the solutions have been: (1) Cut and fit model, proposed by E.L. Thorndike, and described by Ghiselli (1956). This approach does not attempt to reach an optimal solution. It ranks jobs in terms of importance, and places each individual in the most important job for which her expected performance is above some minimal level. (2) The differential prediction approach, which tries to place each individual into the job he is most suitable for. (3) The discriminant function approach, which places individuals into jobs that are filled by similar individuals.

Each of these three proposed solutions has some fault (see Ghiselli, 1956). There exists, however, another approach, which is mentioned by Cascio (1998), and briefly described by Ghiselli (1956). This approach is described in the next paragraph.

Proposed solution

We thought that the structure of our problem makes it best to treat it as a linear programming problem, and solve it as such. The term ‘linear programming’ is taken from the field of operations research. In non-technical terms, linear programming offers optimal solutions to problems in which one tries to maximize some total benefit, or minimize some total damage/loss, under a set of demands, or constraints. In our application, we had a fitness score for every soldier for every possible job. There was a set of constraints, that included quotas: number of soldiers required in every job, rules of the military system (such as which soldier is eligible for which job), and some requirements pertaining to the distribution of soldiers’ “quality” within the various jobs. Our objective was to assign the soldiers to jobs in a way that would maximize the total sum of the fitness scores while not stepping outside the boundaries set by the demands and constraints.

There are two advantages to the linear programming approach: (1) the language of linear programming fits in a straightforward manner into the human resource problem at hand; (2) there exist powerful and effective off-the-shelf algorithms to solve such problems. In our particular application we used SAS proc. LP, but there are other programs in the market.

It should be explicitly stated that the application presented here has two different dimensions. One dimension has to do with the way we computed the fitness scores.
other dimension involves the decision to use a linear programming approach to optimize placement (i.e. to maximize the sum of the fitness scores, subject to constraints). These two dimensions are independent of each other: we could have computed the fitness scores in a different manner, and still used linear programming for optimization of placements, or we could have used other techniques to place the soldiers, based on the fitness scores as we calculated them. Since the present meeting has to do with testing, rather than operations research, the remaining paragraphs will be devoted to the more testing-oriented properties of our work: the fitness scores and some issues of validation.

The fitness scores

In most selection models, including the ones described earlier, calculation of the “fitness scores” uses selection variables that predict job performance. In the United States, for instance, the “fitness scores” upon which admittance into undergraduate studies is based is the candidate’s SAT score. In our own country, it is based on a test similar to SAT plus high school performance. Many potential employers accept or reject candidates by a score composed of test scores, recommendations and interviews.

However, predicted performance, while of major inherent value, is not necessarily the only candidate’s property that can be expressed by his or her fitness score. One simple example is job endurance: we are usually interested in filling positions with workers who will not quit. Endurance is not necessarily correlated with performance while on the job, and is definitely not identical to it. Therefore, one might want to use a composite score that reflects both: performance and endurance. A slightly more complicated example for non-performance-related variables is candidate’s wishes. In a “democratic” situation, the candidate will apply only to jobs she/he is interested in, and there is an option to reject a job offer. But in some situations, such as ours, people can be placed into jobs that are low on their preference list. It should be pointed out that the attempt to take the candidate’s wish into consideration does not necessarily reflect the belief that the higher the wish for the job, the better the expected performance on it is. It can express the decision maker’s values and moral approach, if she believes that the person’s wishes should be taken into consideration while making decisions about him or her. Note that this aspect has some implications for questions of validation. Another example of a variable that is not related to predicted performance, is the importance of the job. If we have a “good” potential worker, who could fit well into any of two different jobs, it stands to reason that he should be placed into the more important job. In sum, the fitness score need not reflect only expected job performance. It can express other issues as well.

In our particular application each soldier’s fitness score to every eligible job included five kinds of variables: (1) Results of differential aptitude tests. Soldiers had been administered four different tests, measuring different abilities, and these different abilities were used to differentially predict performance in the various jobs. For instance: mechanical ability weighed high in calculating fitness scores for mechanical jobs, and clerical ability weighed high for clerical jobs. (2) The second kind of variable was soldier’s wish for the particular jobs. As a rule, the higher a soldier’s wish for a job was, the higher was his fitness score for that job. (3) A third variable reflected soldiers’ evaluations by their commanders. (4) The fourth kind of variable that affected the fitness score was job importance. (5) The fifth variable used in calculating the fitness score was the course of studies the soldier had in high school: For some jobs (but not all), the high
school course of study could be relevant or irrelevant, that is: it could have been expected to enhance job performance, or not to affect it. Of course, other considerations can be taken into account while calculating the score, such as peer rating, bodily measures, etc.

Combining the various variables into a fitness score

The various data for every soldier X job need to be combined into one score. This is the score we’ve been referring to as ‘fitness score’. How does one go about doing it? There are two major approaches to do that: one is empirical in nature, the other – “theoretical” (for lack of a better word). The empirical approach is the typical one taken by selection experts. It requires a criterion (usually, some measure of job performance). Combining of the various data into one composite score is based on regression analysis, with that criterion as dependent variable, and the existing data about the candidates as independent variables. In this approach, weights assigned to the various predictor variables are empirically derived.

In the alternative approach, which is the one we took in this project, the “formula” by which one calculates the fitness score for every soldier X job is done in an interactive process with decision makers, or experts. During such a process, one attempts to extract a set of weights, which presumably exists in the expert’s mind, and which expresses the importance she assigns to the various data. In a way, the situation is similar to that, in which one makes a decision about buying a car: one rates one of the options as “best” for him, based on a host of data, such as cost of the car, its shape, the social image it projects, gasoline consumption, service record, and so on. Apparently, some data are weighed more heavily than others for a given individual, which is why some people buy cars according to cost, and some according to social image. These different weights are not based on regression analyses, nor are they based on predicted performance. They reflect belief systems, preferences, the person’s character, as well as some cultural and environmental influences.

Techniques to extract the weights that decision makers assign to various data have been a major topic of endeavor in the field of MAU (multi-attribute utility. e.g. Edwards 1977; Keeney & Raiffa, 1976). In the project described here, the weights to be used in calculating fitness score were derived in a number of sessions with a particular decision maker. Techniques suggested by MAU experts were used in a somewhat informal manner in these sessions. It should be pointed out that if data, specifically criterion data, are not available, the empirical approach cannot be implemented. Another point to be stressed is that the empirical approach cannot handle situations in which values or job importance are expressed in the composite score. Situations like these pose no problem for the MAU approach we took. When the need arises for a systematic placement system as a decision aid, prior to availability of data, this seems the right approach.

Validation

In regular selection terminology ‘validation’ refers to determining how well the fitness scores predict job performance. This is usually achieved by assessing the correlation coefficient between the predictor and the criterion. However, in our application, such a correlation is not necessarily predicted, and need not necessarily be high. There are two reasons for that: One is that some of the variables used to compute the fitness scores should not correlate with performance. Job importance is one example. The second reason is slightly more complicated, and has to do with the notion of
optimization: we are not trying to place each soldier into the job most appropriate for him. We are trying to optimize the total placement of all soldiers. This means that some soldiers might be placed in a position for which they do not fit too well. In fact, the example given above, in table 2, presents such a situation.

If the regular predictive validity approach to validation is not applicable, how can one validate an application such as the one presented here? One approach will start with the claim that since the system attempts to create the best possible set of placements, then, once it has been implemented, the total organizational performance should improve. Thus, validation of the application will be achieved by examining organizational performance.

The notion that the quality of placement can be tested by comparing organizational performance with – and without – applying the placement scheme sounds attractive, but is unrealistic. Organizational performance, in itself a problematic concept, is heavily affected by factors other than how well employees are assigned to jobs. Market situation as well as a host of other environmental conditions will wash out any effect due to implementation of an optimal placement application. A more realistic approach, and the one we took in our project, involves: (1) Usage of predictive validity procedures to validate those variables in the fitness score that should predict performance, such as test scores; (2) Validation of the weights assigned to the various variables in calculating the test scores, along with the quality of the assignment. We are now engaged in doing this, by creating simulative situations, and comparing placement decisions made by experts, to proposed assignments made by our model.

Limitations and difficulties

Our experience with implementation of the model has shown three main sources of difficulty: (1) Unfeasibility: sometimes the demands cannot be met by supplies (too few soldiers, soldiers lower in quality than required, and so on). In these situations, the model’s response is: “no feasible solution”. Whereas supplies not meeting demands cannot be blamed on the model, it is not simple to detect the source of unfeasibility. (2) There are difficulties in grasping what is “in the expert’s head” when the fitness score is calculated. One gets contradictions among statements, or between estimated weights and actual decisions. To overcome this problem, one might need quite a few iterations with the decision maker. (3) The third limitation is the result of the previous two: in order to apply the model, one needs a well-informed operator. It is impossible to design an application, and assume that the operator will feed in the necessary data, and then just “push the button”.

Conclusions

It is our belief that the approach we took is better than the alternatives presented, it definitely surpasses human computation abilities, and its form fits very well into the structure of some placement situations. However, our specific application needs more work, particularly in validating the model.

References


MEASURING AND PREDICTING CURRENT AND FUTURE NCO PERFORMANCE

Christopher E. Sager, Ph.D.
Principal Staff Scientist
Human Resources Research Organization
66 Canal Center Plaza, Alexandria, VA 22314
csager@humrro.org

Dan J. Putka, Ph.D.
Research Scientist
Human Resources Research Organization
66 Canal Center Plaza, Alexandria, VA 22314
dputka@humrro.org

Deirdre J. Knapp, Ph.D.
Manager Assessment Research and Analysis Program
Human Resources Research Organization
66 Canal Center Plaza, Alexandria, VA 22314
dknapp@humrro.org

INTRODUCTION

One of the central goals of the NCO21 research program was to develop and assess the validity of instruments for use in support of the U.S. Army’s promotion system for noncommissioned officers (NCOs) at pay grade levels E5 and E6. More specifically, this part of the effort was designed to evaluate the validity of instruments developed to predict current and expected future performance.

The Method section of this paper describes (a) the development of criterion scores (i.e., supervisor ratings) for current and expected future performance, (b) the development of predictor scores representing each predictor instrument, and (c) the sample of soldiers involved in the validation data collection. The Results section will describe selected validation analyses and their results. The Discussion section will evaluate each predictor’s potential and address other issues identified during the analyses.

---

34 This paper is part of a symposium titled U.S. Army Non-Commissioned Officer of the Future: Building a Leadership Assessment Tool presented at the 2002 International Military Testing Association Conference in Ottawa (T.S. Heffner, Chair). Substantial portions of it were presented at the 17th Annual Conference of the Society for Industrial Organizational Psychology (Sager, Putka, & Burnfield, 2002). The views, opinions, and/or findings contained in this paper are those of the authors and should not be construed as an official U.S. Department of the Army position, policy, or decision.
METHOD

Criteria

As described in the first presentation (Heffner & Knapp, 2002), two types of supervisor performance ratings criteria were developed: (a) Observed Performance Rating Scales and (b) Expected Future Performance Rating Scales. These two instruments were designed to describe and evaluate E5 and E6 soldiers on requirements relevant to all Army jobs (i.e., Military Occupational Specialties [MOS]).

The Observed Performance Rating Scales were developed in three stages. First, a rating scale was developed for each of 27 performance requirements. Next, the prototype scales including rater instructions and oral rater training were pilot tested on three separate occasions, and then field tested in three separate locations. Two changes resulting from the pilot tests and field test were (a) modification of the rater training and instructions so that supervisor raters could select a “cannot rate” option rather than infer a particular level of subordinate performance on a scale when they had not actually observed it and (b) reduction of the number of scales from 27 to 19 to make the rating task more reasonable. The primary goal of these changes was to maximize the validity and interrater reliability of the observed performance ratings. The 19 scales included performance requirements such as Common Task Knowledge and Skill, Adaptability, and Leadership Skills (for a complete list see Knapp et al., 2002). Each scale consisted of (a) a title of the performance requirement being rated, (b) a one-sentence description of the requirement, and (c) a 7-point scale with separate behavioral anchors for points 1-2, 3-5, and 6-7, respectively.

The Expected Future Performance Rating Scales were also developed in three stages. There was a concern that if scales designed to assess expected future performance took on the same format as the Observed Performance Rating Scales, method variance would result in an artificially high degree of correspondence between scales assessing observed and expected future performance. To minimize this problem, we used themes identified in a future-oriented job analysis (Ford, R. Campbell, J. Campbell, Knapp, & Walker, 2000) to develop six scenarios describing conditions NCOs would be likely to face in the future. These scenarios covered topics like the increasing requirements for technical skill, self-direction, and self-management. Each scenario was between one third and one half of a page long and was followed by a 7-point scale on which the supervisor rated the subordinate’s expected performance effectiveness in the predicted future condition. Similar to the Observed Performance Rating Scales, these Future Performance Scales were pilot-tested and then administered as part of a field test.

The validation data collection involved administering criterion and predictor measures to soldiers and supervisors at seven locations. At the conclusion of the validation data collection, we had observed performance rating scale scores for 1,001 soldiers and expected future performance rating scale scores for 1,012 soldiers. For the Observed Performance Rating Scales, 513 soldiers received ratings from a single
supervisor, 436 soldiers received ratings from two supervisors, and the remaining soldiers received ratings from three or more supervisors. The numbers of supervisor raters per soldier were very similar for the expected future performance ratings. Scale scores were computed for each soldier by averaging the ratings of the supervisors who rated that soldier.

Factor analyses of the Observed Performance Rating Scales were conducted to (a) determine the latent structure underlying these ratings and (b) develop observed performance scores for use in criterion-related validity analyses. Results of these analyses did not show strong support for more than a single factor; therefore, an Observed Performance Composite score was calculated for each soldier based on the mean across the scales. Additionally, a single Expected Future Performance Composite score was calculated for each soldier based on the mean across the six Expected Future Performance Scales. The Observed Performance Composite score interrater reliability estimates for E5 and E6 soldiers were .53 and .59, respectively. The Expected Future Composite score interrater reliability estimates for E5 and E6 soldiers were .40 and .46, respectively.

Predictors

Predictors involved in this validation study are described in Heffner & Knapp (2002). Here, we briefly describe the predictor scale scores that are included in the validation analyses presented here.

The Armed Services Vocational Aptitude Battery (ASVAB) is an operational instrument; the General Technical (GT) score is a composite based on three of its subtests (i.e., Word Knowledge, Paragraph Comprehension, and Arithmetic Reasoning).

In another presentation at this conference Waugh, Sager, & Putka (2002) describe the Situational Judgment Test (SJT) and SJT-X in detail. There is a version of the Situational Judgment Test (SJT) for each pay grade (i.e., E5 and E6 soldiers); each contains some pay grade-specific and some common items. The validation analyses include a single SJT composite score for E5 soldiers and E6 soldiers. The SJT-X was administered to E6 soldiers only; thus no scores are available for E5 soldiers. For E6 soldiers the validation analyses include a single SJT-X composite score.

A semi-structured interview was administered to E5 soldiers only. Each soldier was asked two questions for each of seven KSA constructs relevant to NCO performance (i.e., Adaptability, Self-Management, Effort, Integrity, Supporting Peers, Leadership, and MOS-Specific Knowledge and Skill). Interviewers provided ratings for these seven constructs and two others based on their observations during the interview (i.e., Oral Communication and Military Presence). Because the senior NCO interviewer’s MOS generally did not match the interviewee’s MOS, there was substantial missing data on the MOS-Specific Knowledge and Skill scale. Therefore, the Interview Composite used in all remaining analyses excludes this scale.
Under the current system, candidates for promotion to E5 and E6 pay grades receive points on a Promotion Point Worksheet (PPW). The Simulated PPW Composite used in these analyses is derived from a subset of the areas covered by the full operational PPW. These areas include (a) awards, certificates, and military achievements; (b) military education; (c) civilian education; and (d) military training (i.e., physical fitness and marksmanship). The Simulated PPW Composite served as our best estimate of a proxy for the current NCO promotion system.

The Experiences and Activities Record (ExAct) has 46 self-report items assessing past experience in a variety of activities. Three scores were derived from this measure: (a) computer experience, (b) supervisory experience, and (c) general experience.

The Assessment of Individual Motivation (AIM) is a forced-choice measure that provides scores for six temperament constructs (i.e., Dependability, Adjustment, Work Orientation, Agreeableness, Physical Conditioning, and Leadership; White, 2002).

The Biographical Information Questionnaire (BIQ) is a self-report measure consisting of biodata-like items that provide scores on eight constructs potentially relevant to the Army NCO promotion system (i.e., Hostility to Authority, Manipulativeness, Social Perceptiveness, Social Maturity, Tolerance for Ambiguity, Openness, Leadership, and Interpersonal Skill; Kilcullen, 2002).

Validation Sample

The validation data collection included a total of 1,442 E5 and E6 soldiers (i.e., \(n_{E5} = 885; n_{E6} = 557\)). As described above, performance ratings were collected for about two thirds of these soldiers.

RESULTS

This section focuses on analyses that address the criterion-related validity of each predictor and the incremental validity of each predictor relative to the Simulated PPW Composite (i.e., a partial simulation of the currently operational Promotion Point Worksheet).

Table 1 shows raw and corrected criterion-related validity coefficients for E5 and E6 soldiers for each predictor score relative to the Observed and Expected Future Performance Composite scores. The corrected values were adjusted for range restriction and criterion unreliability. Observed and expected future performance were generally more predictable for E5 than E6 soldiers. The strongest predictors of observed performance for E5 soldiers were AIM Work Orientation, the SJT Composite, AIM Leadership, and BIQ Leadership. The strongest predictors of observed performance for E6 soldiers were the SJT Composite, ASVAB GT, the SJT-X Composite, and BIQ Interpersonal Skill. The strongest predictors of expected future performance for E5 soldiers were AIM Work Orientation, AIM Leadership, BIQ Leadership, and the SJT Composite. The strongest predictors of expected future performance for E6 soldiers were
the SJT Composite, the SJT-X Composite, ExAct Computer Experience, and BIQ Interpersonal Skill.
**Table 1. Correlations between Predictor and Criterion Scores by Pay Grade**

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Observed Performance Composite</th>
<th>Expected Future Performance Composite</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw</td>
<td>Corrected</td>
</tr>
<tr>
<td></td>
<td>E5</td>
<td>E6</td>
</tr>
<tr>
<td>PPW Composite</td>
<td>.19</td>
<td>.09</td>
</tr>
<tr>
<td>ASVAB GT Score</td>
<td>.08</td>
<td>.11</td>
</tr>
<tr>
<td>SJT Composite</td>
<td>.23</td>
<td>.16</td>
</tr>
<tr>
<td>SJT-X Composite</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interview Composite</td>
<td>.17</td>
<td>.14</td>
</tr>
<tr>
<td>ExAct Computer Experience</td>
<td>.09</td>
<td>.07</td>
</tr>
<tr>
<td>ExAct Supervisory Experience</td>
<td>.08</td>
<td>-.02</td>
</tr>
<tr>
<td>ExAct General Experience</td>
<td>.13</td>
<td>.07</td>
</tr>
<tr>
<td>AIM Dependability</td>
<td>.11</td>
<td>-.01</td>
</tr>
<tr>
<td>AIM Adjustment</td>
<td>.06</td>
<td>.07</td>
</tr>
<tr>
<td>AIM Work Orientation</td>
<td>.28</td>
<td>.09</td>
</tr>
<tr>
<td>AIM Agreeableness</td>
<td>.01</td>
<td>-.01</td>
</tr>
<tr>
<td>AIM Physical Conditioning</td>
<td>.11</td>
<td>.02</td>
</tr>
<tr>
<td>AIM Leadership</td>
<td>.22</td>
<td>.06</td>
</tr>
<tr>
<td>BIQ Hostility to Authority</td>
<td>-.06</td>
<td>-.13</td>
</tr>
<tr>
<td>BIQ Manipulativeness</td>
<td>-.08</td>
<td>-.10</td>
</tr>
<tr>
<td>BIQ Social Perceptiveness</td>
<td>.15</td>
<td>-.01</td>
</tr>
<tr>
<td>BIQ Social Maturity</td>
<td>.06</td>
<td>.05</td>
</tr>
<tr>
<td>BIQ Tolerance for Ambiguity</td>
<td>.14</td>
<td>.04</td>
</tr>
<tr>
<td>BIQ Openness</td>
<td>.05</td>
<td>-.06</td>
</tr>
<tr>
<td>BIQ Leadership</td>
<td>.25</td>
<td>.04</td>
</tr>
<tr>
<td>BIQ Interpersonal Skill</td>
<td>.11</td>
<td>.14</td>
</tr>
</tbody>
</table>

*Notes. n_{E5} = 471-613; n_{E6} = 341-399. ‘Corrected’ correlations were corrected for criterion unreliability and range restriction on the predictor. Statistically significant correlations are bolded, p < .05 (one-tailed).*

Table 2 shows raw and corrected incremental validity coefficients for each predictor score beyond the Simulated PPW Composite. The purpose of these values is to estimate the extent to which each predictor could improve the prediction of performance beyond the currently operational PPW. Predictors with strong incremental validity for E5 soldier observed performance were AIM Work Orientation, BIQ Leadership, and the SJT Composite. Predictors with relatively strong incremental validity for E6 observed performance were the SJT Composite and SJT-X Composite. Predictors with strong incremental validity for E5 expected future performance were AIM Work Orientation, BIQ Leadership, and AIM Leadership. Predictors with relatively strong incremental...
validity for E6 expected future performance were the SJT Composite, SJT-X Composite, and AIM Adjustment.

Table 2. Incremental Validity Estimates of Predictor Score Composites by Pay Grade

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Raw Observed Performance Composite</th>
<th>Raw Expected Future Performance Composite</th>
<th>Corrected Observed Performance Composite</th>
<th>Corrected Expected Future Performance Composite</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E5</td>
<td>E6</td>
<td>E5</td>
<td>E6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASVAB GT Score</td>
<td>.01</td>
<td>.06</td>
<td>.01</td>
<td>.05</td>
</tr>
<tr>
<td>SJT Composite</td>
<td>.09</td>
<td>.10</td>
<td>.10</td>
<td>.09</td>
</tr>
<tr>
<td>SJT-X Composite</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interview Composite</td>
<td>.05</td>
<td></td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ExAct Computer Experience</td>
<td>.01</td>
<td>.02</td>
<td>.01</td>
<td>.05</td>
</tr>
<tr>
<td>ExAct Supervisory Experience</td>
<td>.01</td>
<td>.01</td>
<td>.03</td>
<td>.00</td>
</tr>
<tr>
<td>ExAct General Experience</td>
<td>.01</td>
<td>.01</td>
<td>.03</td>
<td>.01</td>
</tr>
<tr>
<td>AIM Dependability</td>
<td>.01</td>
<td>.00</td>
<td>.04</td>
<td>.00</td>
</tr>
<tr>
<td>AIM Adjustment</td>
<td>.01</td>
<td>.02</td>
<td>.01</td>
<td>.06</td>
</tr>
<tr>
<td>AIM Work Orientation</td>
<td>.13</td>
<td>.03</td>
<td>.18</td>
<td>.03</td>
</tr>
<tr>
<td>AIM Agreeableness</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>AIM Physical Conditioning</td>
<td>.02</td>
<td>.00</td>
<td>.03</td>
<td>.01</td>
</tr>
<tr>
<td>AIM Leadership</td>
<td>.08</td>
<td>.02</td>
<td>.15</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIQ Hostility to Authority</td>
<td>.01</td>
<td>.06</td>
<td>.02</td>
<td>.03</td>
</tr>
<tr>
<td>BIQ Manipulativeness</td>
<td>.01</td>
<td>.04</td>
<td>.02</td>
<td>.04</td>
</tr>
<tr>
<td>BIQ Social Perceptiveness</td>
<td>.05</td>
<td>.00</td>
<td>.09</td>
<td>.00</td>
</tr>
<tr>
<td>BIQ Social Maturity</td>
<td>.00</td>
<td>.02</td>
<td>.00</td>
<td>.02</td>
</tr>
<tr>
<td>BIQ Tolerance for Ambiguity</td>
<td>.04</td>
<td>.01</td>
<td>.06</td>
<td>.02</td>
</tr>
<tr>
<td>BIQ Openness</td>
<td>.01</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
</tr>
<tr>
<td>BIQ Leadership</td>
<td>.11</td>
<td>.00</td>
<td>.18</td>
<td>.01</td>
</tr>
<tr>
<td>BIQ Interpersonal Skill</td>
<td>.03</td>
<td>.07</td>
<td>.03</td>
<td>.06</td>
</tr>
</tbody>
</table>

Notes. \( n_{E5} = 432-435; n_{E6} = 296-300 \). “Corrected” coefficients were corrected for criterion unreliability and multivariate range restriction. Statistically significant correlations are bolded, \( p < .05 \) (one-tailed).

DISCUSSION

The analyses presented here indicate that several predictors show potential as additions to the US. Army’s promotion process to pay grades E5 and E6 (i.e., junior NCOs). They also reveal some interesting issues about the predictability of performance across pay grades.
ASVAB GT showed better criterion-related validity and incremental validity over the Simulated PPW Composite for E6 soldiers than for E5 soldiers. It is important to note, however, that in other Army research, ASVAB scores have shown substantially better criterion-related validity when predicting non-ratings criteria such as scores on job simulations and job knowledge tests (Campbell & Knapp, 2001).

The SJT showed strong criterion-related validity and incremental validity over the Simulated PPW Composite for E5 soldiers; however, the evidence was more moderate for E6 soldiers. For a relatively short instrument (i.e., 3 items), the SJT-X showed impressive criterion-related and incremental validity.

It is important to note that this effort’s simulation of the PPW is different from the operational PPW in two important ways: (a) we collected this information via a self-report instrument instead of from archival records and (b) we did not include two elements of the operational PPW (i.e., points from the recommending commander and a promotion board appearance). However, our simulated composite included most of the PPW points.

The three ExAct scales showed good criterion-related validity estimates; however, they did not show strong incremental validity relative to the Simulated PPW Composite. This is not a surprising result given that both measures cover relatively similar domains of experience and training.

Scales from the AIM and the BIQ showed strong criterion-related for validity and incremental validity over the Simulated PPW Composite for E5 soldiers and some criterion-related validity for E6 soldiers. An AIM scale even showed some incremental validity for E6 soldiers. Results for these instruments are discussed in more detail in White (2002) and Kilcullen (2002).

One of the most notable results of this research is the substantial difference in criterion-related validity estimates for E5 soldiers versus E6 soldiers. None of the numerous proposed explanations for this difference have been supported by analyses (e.g., differences in the composition of the E5 and E6 soldier samples on potentially influential characteristics like type of MOS or gender).

Finally, past U.S. Army research (e.g., Campbell & Knapp, 2001) has considered supervisor ratings as primarily measures of “will-do” performance. That is, a supervisor’s ratings are substantially influenced by perceptions about the subordinate’s level of effort, whereas job simulations and knowledge tests are considered measures of “can-do” performance. The idea is that a “can-do” measure assesses the extent to which the subordinate has the knowledge and/or skill necessary to perform the work. Any conclusions derived from these analyses need to recognize that our criteria were supervisor ratings. Can-do measures of performance may have had different patterns of relationships with the predictors.
REFERENCES


ABSTRACT

Low response rates increasingly affect military and civilian surveys. The Navy Equal Opportunity/Sexual Harassment (NEOSH) Survey obtained a response rate of 60% in 1989; by 1999, the rate was 30%. This study was conducted to determine reasons for survey non-response from U.S. Navy personnel members originally selected for participation in the 1999 NEOSH survey. This study also gathered potential solutions to the non-response issue. The results indicated a belief that surveys have no impact was cited as a common reason why they aren’t being returned. Suggestions for increasing response rates most commonly included providing feedback, increasing command involvement, shortening survey length, and providing incentives. When the responses of those who said they completed the NEOSH survey were compared to those who indicated they did not, the responses of the two groups were similar for attitudinal items. However, completers endorsed factual items assessing attendance at EO-related training more than noncompleters.

INTRODUCTION

Lower response rates are an issue that increasingly affect both military and civilian-sector surveys. Reviews of academic and organizational surveys have documented a gradual decline in response rates, especially those that are administered by mail (Baruch, 1999; Rogelberg, Luong, Sederberg, & Crystol, 2000). In a review of academic studies published in leading management journals, Baruch (1999) reported an average response rate of 64% in 1975, 56% in 1985 and 48% in 1995.

This trend towards lowered response rates in academic and organizational surveys in the civilian sector has also been obtained on U.S. military personnel surveys (Knouse, 2002). In the Navy, Edwards, Rosenfeld, Booth-Kewley, and Thomas (1996) noted that response rates on Navy mailout surveys decreased during the early 1990s. This downward trend has continued. The Navy-wide Personnel Survey, had a 52% response rate in 1990, a 45% response rate in 1996 and a 33% response rate in 2000. The NEOSH response rates have similarly decreased from 60% in 1989 to 40% in 1995 to 30% for the 1999/2000 survey. Lamerson (2001) reports that response rates on mail-out surveys administered in member nations of The Technical Cooperation Program (United States, United Kingdom, Canada, Australia, and New Zealand) declined during the 1990s and continue to decline. In sum, lower response rates appear to be a universal trend in mail-out surveys occurring in civilian and military settings in both the United States and other countries.

35 The opinions expressed are those of the authors. They are not official and do not represent the views of the U.S. Navy Department. The authors appreciate the assistance of the project sponsor, CDR Leanne Braddock.

36 Rorie Harris and Regina Hindelang are in the Psychology Department at the University of Memphis, Memphis, TN.
These lower response rates are of concern for a number of reasons. Lower response rates result in less survey data, potentially limiting the generalizability of the results. This may create the need for larger sample sizes on future surveys which will increase the already substantial costs incurred by large-scale mailout surveys. Low response rates also may heighten the impact of nonresponse bias, which occurs when those who do not respond to the survey are systematically different on key variables from those who respond (Rogelberg & Luong, 1998). Low response rates can impact the credibility of the survey since survey sponsors often use response rates as an index of a survey's quality (Edwards, Thomas, Rosenfeld, & Booth-Kewley, 1997).

As response rates have declined, attention has focused on reasons for why the decline is occurring. Among the reasons offered have been oversurveying of respondents who receive excessive numbers of surveys (Rogelberg, et al., 2000), lower motivation to respond especially in large organizations due to lack of knowledge or commitment to the survey topic (Knouse, 2002; Tomaskovic-Devey, Leiter, & Thompson, 1994), less time to respond due to increased workloads, and lack of perceived benefit to respondents from responding (Baruch, 1999). However, there has been little empirical data in the civilian sector and none that we are aware of in military survey research that supports any of these reasons for non-response or offers suggestions from respondents for improving response rates. Indeed, it is impossible to know precisely what might be causing lowered responses rates and what steps would effectively increase response rates without directly contacting those who were sent the survey. To address this issue, a study was conducted to determine reasons for survey non-response from U.S. Navy personnel members originally selected for participation in the 1999 NEOSH survey. Potential solutions to survey non-response were also assessed.

METHOD

A random subset of 4,074 active-duty Navy personnel from the original 1999 NEOSH sample was sent a short survey which solicited their views on issues related to lower response rates and possible solutions for increasing response rates. The survey also contained several questions from the NEOSH survey so that the responses of NEOSH completers and NEOSH non-completers could be compared. Phone cards were included in the survey packets and those who had not completed the NEOSH survey were asked to call in and participate in a more in-depth telephone interview.

RESULTS

Survey

Surveys were returned by 1,368 respondents. After adjusting for the 596 surveys that were returned as undeliverable and the 31 surveys that were incomplete, the response rate was 39%, which is 9 percentage points higher than the response rate obtained on the 1999 NEOSH survey. About half (53%) indicated that they completed the 1999 NEOSH survey, 21% reported that they didn’t complete the survey, and another 25% reported that they didn’t know if they completed it. Answers to open-ended questions on the completed surveys were typed into a database, content analyzed by grouping similar comments into key areas, and analyzed using SPSS. Respondents were asked the number of Navy surveys they had received in the past year, explanations for declining response rates, suggestions on how to improve response rates, and several NEOSH Survey items.
When asked to respond to the question “Why have response rates on Navy surveys declined?” the most frequently occurring responses were “No changes result/Response doesn’t matter”, “Apathy/Laziness/Tired of surveys”, “Survey length/Time consuming”, “Too busy/Low priority”, and “Too many surveys” (see Fig. 1). These top five accounted for two-thirds of the responses.

Respondents were also asked what steps should be taken to increase responses to Navy surveys. As shown in Fig. 2, the top five responses were “Provide feedback/Publish results”, “Command involvement”, “Make surveys shorter/More concise”, “Provide incentives”, “Provide incentives”, and “Put survey online”. These top five accounted for over half of all responses.

Since it is commonly believed that Navy personnel do not complete surveys...
because they receive too many surveys, respondents were asked how many Navy surveys they received in the past year. On average, respondents reported receiving an average of 2.00 Navy surveys in the past year, with officers receiving more surveys than enlisted personnel did (see Table 1). These results do not support the perception that Navy personnel are over-surveyed.

<table>
<thead>
<tr>
<th>Paygrade</th>
<th>Mean</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2-E4</td>
<td>1.42</td>
<td>146</td>
</tr>
<tr>
<td>E5-E6</td>
<td>1.79</td>
<td>383</td>
</tr>
<tr>
<td>E7-E9</td>
<td>1.85</td>
<td>122</td>
</tr>
<tr>
<td>W2-W4</td>
<td>2.25</td>
<td>16</td>
</tr>
<tr>
<td>O1-O2</td>
<td>2.33</td>
<td>27</td>
</tr>
<tr>
<td>O3 and above</td>
<td>2.35</td>
<td>498</td>
</tr>
<tr>
<td>Total</td>
<td>2.00</td>
<td>1192</td>
</tr>
</tbody>
</table>

To test whether those who completed the NEOSH Survey differ from those who did not, the responses of completers and non-completers, and those who didn't know if they completed the survey were compared. Respondents were asked several factual (e.g., whether they attended sexual harassment, and equal opportunity training in the past year) and attitudinal (e.g. satisfaction with the Navy) NEOSH Survey items. As can be seen in Table 2, completers and noncompleters had similar responses on attitudinal items but differed in their responses to factual items. On factual items, completers were more likely than noncompleters to report attending EO, SH and Fraternization training. The responses of those who did not know whether they completed the NEOSH Survey were similar to the responses of completers.

<table>
<thead>
<tr>
<th>Factual Items</th>
<th>Completers</th>
<th>Noncompleters</th>
<th>Don’t Know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Received EO training at this command</td>
<td>81%</td>
<td>70%</td>
<td>82%</td>
</tr>
<tr>
<td>Attended Fraternization training at this command</td>
<td>85%</td>
<td>79%</td>
<td>85%</td>
</tr>
<tr>
<td>Received Sexual Harassment training</td>
<td>87%</td>
<td>81%</td>
<td>88%</td>
</tr>
<tr>
<td>Attitudinal Items</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Would recommend the Navy</td>
<td>79%</td>
<td>77%</td>
<td>76%</td>
</tr>
<tr>
<td>Satisfaction with Navy</td>
<td>80%</td>
<td>79%</td>
<td>80%</td>
</tr>
</tbody>
</table>

Note: N= 707 completers, 279 noncompleters, 338 DK; a=significant difference from noncompleters
Phone Interviews

Respondents who had not completed the 1999 NEOSH survey were asked to complete a phone interview. One hundred and ten respondents were subsequently interviewed. Of this number, 25 reported that they had not returned the 1999 NEOSH survey, 28 stated that they completed the survey, and 57 said that they don’t know if they completed the NEOSH survey.

Phone respondents were asked to provide reasons why fewer people are returning Navy surveys. Their responses were similar to those received on the short survey. The top 5 reasons were “Too lengthy/Time consuming”, “Don’t see as important”/Don’t take seriously”, “Apathy/Laziness”, “No feedback/Inadequate feedback”, and “Too busy/Low priority”. Respondents also provided similar suggestions on how to get more people to complete and return Navy surveys to those obtained on the survey.

Only 28% of the phone interviewees agreed that Navy people receive too many surveys. Seventy percent reported that they had never heard or been told of the results of a Navy survey. This may explain why many suggested providing survey results to respondents as a way of increasing response rates. Phone interviewees were generally neutral or positive about the Navy surveys they receive and 81% reported that if they received a NEOSH survey next year they would complete it. Most believe the survey confidentiality guarantees stated on surveys and more than half reported that they thought the Navy takes survey results seriously.

DISCUSSION

Although there has been much speculation about the reasons for nonresponse, this study provides the first empirical evidence from actual Navy survey respondents. The responses of over 1,300 individuals who were in the NEOSH 1999 sample indicate that, contrary to popular belief, Navy personnel are not oversurveyed (at least by Navy surveys). Rather, a belief that surveys have no impact was cited as a common reason why they aren’t being returned. This coupled with survey length and a general apathy over the survey process suggest that Navy surveys need to better engage respondents, be short and concise as possible, and be utilized by policymakers in a manner that is clear to respondents. Suggestions for increasing response rates most commonly included providing feedback, increasing command involvement, shortening survey length, and providing incentives. Thus, providing feedback about the results, better utilization of the results, and making surveys shorter will likely increase response rates. Incentives, are more problematic to implement, since incentives for participation in military surveys are currently prohibited—the $5 phonecard was permitted on an experimental basis since it also allowed respondents to actually contact the researchers.

When the responses of those who said they completed the NEOSH survey were compared to those who indicated they did not complete the survey, the responses of the two groups were similar for attitudinal items. However, for factual items assessing attendance at EO-related trainings, noncompleters had generally lower responses than completers. Noncompleters compared to completers were less likely to say they attended trainings; while responding similarly to attitudinal items. This finding suggests that survey completers may be those who are more organizationally compliant than non-completers. It also provides indirect support that on the NEOSH attitudinal items, the
findings are not being systematically affected by nonresponse bias since completers and noncompleters responded similarly.

A number of the findings of the present study were implemented on the 2002 NEOSH Survey which was sent to a random sample of Navy personnel in September 2002. The survey was shortened by about 1/3, several Navy news notices about the survey were released encouraging responses. Navy leadership also expressed an intention to potential respondents to provide feedback about the results. Response rates on the 2002 NEOSH will be used to assess whether these steps impact survey returns.

REFERENCES
TRAINING COMMUNICATION COMPETENCE FOR MULTINATIONAL TEAMS

Sharon L. Riedel
U.S. Army Research Institute, Leader Development Research Unit
851 McClellan Ave, Fort Leavenworth, Kansas USA 66027
Sharon.Riedel@leavenworth.army.mil

Angela Karrasch
U.S. Army Research Institute, Armored Forces Research Unit
Bldg 2423 Morande St., Fort Knox, Kentucky, USA 40121
Angela.Karrasch@Knox.army.mil

ABSTRACT

Communication is critical in any team setting but particularly with multinational military teams. Recent events highlight the realization that more military multinational collaboration will be necessary. However, the scant amount of research that exists on multinational teams indicates that communication is one of the most pressing difficulties affecting team efficiency and effectiveness. But communication competence can be developed through education and guided practice.

This paper presents a framework and methodology for training multinational team (MNT) communication competence. Requirements for the training were derived from focus groups conducted at Headquarters (HQ) Stabilization Force (SFOR) at Camp Butmir, Bosnia-Herzegovina. Results from focus group data are presented and discussed. The training methodology is then described. The methodology uses an instruction with small group discussion format. Participants’ experiences and problems with communication in MNTs are elicited and discussed. The Instructor guides the participants to a discussion of the major skills and attitudes needed for MNT communication competence. Throughout the discussion, the facilitator highlights and elaborates teaching points related to cultural communication styles and communication strategies. Results of a pilot study of the training methodology, also conducted with MNTs at HQ SFOR, indicate that participants saw the training as needed and useful. They especially liked the focused discussion format for this type of training.

INTRODUCTION

The United States Armed Forces is probably one of the best trained forces in the world - for fighting. However, because of globalization and related factors, the U.S. military troops are performing missions that, for the most part, do not require fighting. The face of war is changing and our troops are now joining with coalition forces to conduct peacekeeping, humanitarian aid, disaster relief, and terrorism prevention and response. These changes require many individual adjustments on the part of the soldiers
— in terms of both attitudes and skills. As one American officer in Bosnia recently put it, “I signed up to fight. I was trained to fight. I want to fight. What am I doing here?”

At a higher level, adjustments in preparations for military operations other than war and multinational teaming must be made. Little research has addressed how multinational teams function and the unique difficulties they face. Questions such as the following are just beginning to be addressed: What unique skills, knowledge and abilities do leaders need to lead multinational teams? How does culture affect the ways different nationalities approach and solve problems. What are the barriers to effective performance in multinational teams?

This paper describes a study to investigate the barriers to effective performance in multinational teams and the training that was developed to address one of the barriers that emerged — i.e., communication. The communication framework that was developed to guide the training program of instruction will be described. Finally, the results of a pilot study to evaluate the training will be described and future plans discussed.

BACKGROUND

This project is part of an Army Research Institute and Army Research Laboratory collaborative research at Headquarters (HQ) Stabilization Force (SFOR), Camp Butmir, Sarajevo. The purpose of this umbrella project is to develop and test methods to better understand the impact of cultural differences on cognitive aspects of teamwork at SFOR. Project scientists traveled to Bosnia-Herzegovina in April and June 2002 to document the lessons learned by MNTs working at the HQ SFOR at Camp Butmir and to explore the barriers to effective performance in multinational teams.

The mission of SFOR includes deterring hostilities and stabilizing the peace, and contributing to a secure environment by providing a continued military presence. HQ SFOR contains about 500 NATO and Non-NATO troops from over 20 different nations. The majority of team members at SFOR HQ are of field grade rank or higher. The official language at SFOR is English. This means that a high proportion of officers there are not communicating in their native language.

TRAINING REQUIREMENT

The requirement for communication training with MNTs came from focus groups conducted in April 2002 at HQ SFOR in Bosnia-Herzegovina. The purpose of the focus groups was to elicit lessons learned by MNTs and barriers to effective MNT performance. Two focus groups were conducted. One group consisted of eight U.S. Army officers, including seven majors and one Lieutenant Colonel. The second group consisted of ten officers from a variety of nations, including United Kingdom, Turkey, Germany, Netherlands, Italy, Spain, Poland, Argentina, and France. Eight Majors and two Lieutenant Colonels were included. An American only group and a mixed

“People don’t understand each other and don’t even know they don’t understand what others are thinking. This problem is very common here. You think you are getting your message through and you’re not…Not being able to communicate [here] is a big problem.”

Spanish Officer, HQ SFOR, Bosnia-Herzegovina, April, 2002
nationality group, without Americans, were selected (1) because it was thought the mixed group would speak more freely without Americans, who are the dominant presence at HQ SFOR, and (2) to examine differences between Americans and other nationalities in perceptions of barriers to team functioning.

A content analysis showed that the Foreign Only group thought that communication is easily the number one challenge to team work in MNTs. One officer said his group felt like they were the Forest Gump of SFOR - they had lots to contribute, but language was a big barrier in making that contribution. Some thought that the practice of giving tasks to native speakers (i.e., speakers whose first language is English), simply because it was easier, hurt the team spirit of MNTs.

Communication in MNTs is degraded by a number of factors. First, just speech mechanics can make communication difficult for non-native speakers. For example, speaking too fast or speaking in an English or American regional accent can make the speaker hard to understand. One non-native speaker said that even at his SFOR Newcomers’ Orientation, the native speaker presenter had a southern accent and he was able to understand only half of what she was saying. Cultural factors can also get in the way of communication, such as the use of acronyms, slang, or humor. Sometimes officers hesitate to ask questions when they don’t understand because in their culture it is rude to ask questions or because one doesn’t question one’s supervisor. Several mentioned the stress that is created by the requirement to communicate all the time in another language. One said, “When we are thinking and listening and speaking in our second language, it takes a tremendous amount of focus and concentration to do that. And we tire a lot faster because of it.” Another said that understanding the language uses 50% of his thinking capacity, leaving only half of his brain power to do his tasks. Participants felt the supervisor plays a critical role in managing the communication problem – he sets the tone of patience and respect for non-native speakers. All agreed that patience and tolerance are the key to the communication problem.

The American group, on the other hand, did not see language as the major problem that non-native speakers did. One participant did say that the key to building and keeping trust was working through the language barrier and letting others know they are contributing. One solution frequently offered by the Americans for the “language problem” was that other countries should send only fluent English speakers to HQ SFOR. One said that there are European “quirks” that require adaptation – hour long coffee breaks in the morning and afternoon, slower pace of work, dinner at 8, and the approach to alcohol consumption.

The groups discussed other barriers to MNT performance in addition to communication, but these are outside the scope of this paper and will not be addressed.

Based on the results of the focus group data, it was concluded that communication appears to be a major difficulty in MNTs. If we could improve the communication skills of MNT members, we might also improve the effectiveness and efficiency of MNTs.

DESCRIPTION OF TRAINING

The objectives of this training are to increase participants’ awareness (1) of possible communication difficulties experienced by MNTs; (2) that the responsibility for communication lies not only with the speaker, but also with the listener, and team leader,
and (3) of strategies for communication competence in MNTs. The training focuses on what the listener, speaker and team leader can do to improve communication in MNT activities.

The training method uses a small group interactive discussion lead by a Trainer. The group discussion format engages people and makes the training relevant to their experiences and the problems they’ve had. The Trainer asks for an example of miscommunication in MNTs they’ve been in and then leads the group in an analysis of the miscommunication. During the discussion, the Trainer highlights, elaborates, and adds to the strategies that have been brought up by the group. These are mini lectures that are inserted throughout the group discussion.

Two aspects of training approach should be pointed out. First, the approach addresses general communication skills rather than teaching information about specific cultures. It would be prohibitively time consuming and costly to educate each SFOR soldier in the culture related communication behaviors of over 20 SFOR participating nations. An alternative is to look at skills for identifying and dealing with cultural differences in communication as they occur. This approach is consistent with a shift in training philosophy from traditional methods that teach what-to-think to training that addresses how-to-think (Klein & Pierce, 2001).

Second, the training format combines presentation of information by the Instructor with small group interactive discussion where the participants relate their experiences to what the Instructor is presenting. Most adult courses involve a lecture portion where a teacher talks and learners take notes – one of the worst arrangements for the promotion of in-depth understanding (Halpern, 2002). The frequent use of real life examples helps students recognize the principles to be learned when they encounter them in different contexts. This is especially true when a wide variety of examples are used and informative feedback is provided. Guided learning or learning with an experiential component is the primary mode of on-the-job learning. Skills and knowledge develop best when a combination of formal and informal learning experiences are used (Grolnic, 2000). The training described here capitalizes on these learning principles.

Fig.1 shows factors that influence communication effectiveness in MNTs. The bullets in each box represent examples and do not make up a complete list. The training will address those boxes that are bolded, i.e. attitudes, cultural conventions, and motivation. Environmental conditions – workload, stress, and fatigue – certainly affect ease of communication. But as with language skills and organizational climate, they are not easily changed.

This training deals with attitude, cultural conventions, motivation and the roles of the speaker, listener and team leader. The team leader sets communication climate of the team. He is the role model. He sets the tone. If he is patient and respectful of non-native speakers, others will be also. If he encourages non-native speakers to ask questions if they don’t understand, others will also. If he asks someone to speak more slowly if needed, others will also.

Attitudes are an important driver of our behaviors. The need for patience and acceptance in MNTS was frequently cited in our focus groups. Another important attitude is to be alert to possible communication problems, both when you’re the speaker, listener and team leader. Being adaptable is important, that is, taking the perspective of the listener when you’re speaking, or the speaker when you’re listening. If you’re the speaker
speaking to a non-native speaker, speak slowly and don’t use acronyms and abbreviations. If you’re the team leader, make sure all perspectives are considered.

Figure 1. Factors that influence communication effectiveness in multinational teams.

Being motivated to understand and to be understood is a big factor in communication effectiveness. Cultural conventions also influence communication. For example, in some cultures, it’s not appropriate to ask questions if you don’t understand. Or there may be the feeling that asking questions will show one’s ignorance. Some cultures are more task oriented in their communication and others are more attuned to the social relationships. The relationship oriented team member may spend too much time on relationships and not on tasks. If they’re task oriented they may inadvertently offend if they do not take the time to learn about other cultures.

In general, the major communication strategies that will be addressed include:

- Be a responsible listener, speaker and team leader
- Be alert for communication problems
- Check that you’ve been understood; check that you understand; get clarification; ask questions
- Consider your audience – if necessary, speak slowly; avoid acronyms and slang; be patient
- Take the other person’s point of view
- Attend to non-verbal communication
- Be self aware

PILOT TESTING OF TRAINING
Procedures and participants

The training was pilot tested in June 2002 at Camp Butmir. Four groups were trained. Each group consisted of a mixture 6 – 8 male, native and non-native speakers, for a total sample size of 28. Thirty seven percent were Majors, 32% Lieutenant Colonels, 3% Colonels, 20% Captains, 5% Lieutenants and 3% Senior NCOs. Training lasted two hours. After the training, participants completed an evaluation questionnaire.

Reactions to the Training

The evaluation questionnaire scale ranged from 1 (strongly disagree) to 7 (strongly agree). The results are shown below.

<table>
<thead>
<tr>
<th>Questionnaire Item</th>
<th>Percent agree to strongly agree (scores 5-7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The training was relevant for my work at HQ SFOR</td>
<td>89</td>
</tr>
<tr>
<td>The discussion on communication contained material that was new or of interest to me</td>
<td>64</td>
</tr>
<tr>
<td>The discussion on communication contained information that will be useful</td>
<td>86</td>
</tr>
<tr>
<td>The discussion on communication is an appropriate method for preparing new members of the SFOR HQ</td>
<td>93</td>
</tr>
</tbody>
</table>

Summary of open ended question responses:

1. What aspects of the discussion did you find useful?
The most frequent response mentioned the open forum of the training with the opportunity to share concerns and identify possible conflict areas. The second most frequently mentioned response concerned hearing the views of native speakers and their attitudes toward non-native speaker communication problems.

2. Describe any aspects of communication that were not addressed in this discussion, but that you believe would be helpful.
Some items mentioned included do’s and don’t for VTC and meetings, and written communication.

3. How can this training be improved?
By far the most frequent comment was that this training should be included in the Newcomers’ briefing or done once a month in an informal setting. The second most frequently mentioned comment was that the non-native speakers needed to be more vocal in the training; the English speakers dominated the session. Some suggested having separate sessions for native and non-native speakers. Another frequently mentioned recommendation was to have slides to help participants remember the main points.

4. Please comment on any cultural differences you have personally encountered that impacted your ability to work as an SFOR HQ team member.
The “language skill problem” was the most frequently mentioned concern – being cited by nine participants. A problem mentioned by two participants was that the different perceptions of rank by certain nations may lead to misunderstandings.

STUDY CONCLUSIONS AND FOLLOW-ON

The results of the focus groups and the pilot study indicated a need for training in communication skills for MNTs. Based on these results, Lt.G. John Sylvester, Commander of SFOR, directed that this training be implemented at the SFOR Newcomers Orientation in October 2002. At this time, SFOR Newcomers Trainers will also be trained to conduct this communication training in future Newcomer’s Orientations. Three products will be produced as part of the training instruction:

- A set of communication strategies for the speaker, listener, and team leader;
- A Trainer’s Manual describing the communication strategies and techniques for facilitating small group discussions
- A Rules of Communication Card to remind MNT members of useful communication strategies

DISCUSSION

This project began as an investigation of the cultural factors that impact performance of MNTs. The one factor that we heard repeatedly, and that seemed to almost overwhelm other factors, was the difficulty non-native speakers had just communicating. Some didn’t understand half of each morning’s briefing. Briefers spoke too fast, used acronyms, and spoke in dialects. Having to continually communicate in another language was a source of cumulative stress for many non-native speakers. If they didn’t understand the first time, often the task was simply given to a native speaker, who could understand. It became clear that if we could simply improve the communication skills of the members of MNTs, we might also be able to impact performance.

This training described here uses simple principles of psychology. It is not costly in either time or dollars. It appears to be training that engages participants. The training is still under development and it remains to be seen if it will have an impact on MNT functioning.

It is very possible that with continued globalization, with the number of multinational corporations multiplying, with military operations increasingly using coalition forces – that MNT communication will not be a problem in the future. English may be the second language of the world. But in the mean time, communication within MNTs can present problems and needs to be addressed. The training proposed here is one small step in that direction.

REFERENCES


INTRODUCTION

Over the past few years, a variety of new approaches to communication have been introduced. Today, technologies such as email, wireless text messaging, discussion groups, computer-video conferencing, and online chat rooms are in use by an increasing number of people in industrialized nations. These new media offer advantages to the average person in lower cost, higher speed and increased accessibility to communication with others that may have been difficult or costly to reach with other more traditional approaches to communication. However, while each of these new methods offer an advantage over the traditional forms of communication, they increasingly separate a person from the typical characteristics of face-to-face communication. These characteristics include access to both the verbal and non-verbal content conveyed in the communication, as well as the actual presence of another person during the communication.

For military researchers, one of the primary questions we are faced with is how if at all we might be able to exploit these new methods of communication to advance research with our military personnel. Central to this consideration are questions about whether personnel will respond to new media, and have access to the necessary technology, and will provide responses of equivalent quantity/quality in comparison to traditional methods of data collection. If these new methods can be adapted and/or exploited in such a manner as to facilitate research that is cost-effective and successful in obtaining the needed data, they will be embraced. In the world of research, online methods are becoming increasingly popular as a means of collecting data for market research, public-opinion surveys, and hosting discussions of political and social issues.

In particular, the chat room appears to be a promising new approach to conduct focus groups. Utilizing an online chat room approach would offer many benefits over the traditional face-to-face focus group method. For instance, traditional focus groups require that either the moderator(s) travel to the participants or the participants travel to the moderator(s), whereas with online focus groups there are no costs associated with travel or lodging. In addition, the online focus groups offer a degree of automation not found in traditional focus groups that can reduce time investment by research professionals. The automatic capture of group transactions by the computer software eliminates the need for manual transcription and allows for possible automation of the response coding process. Finally, online focus groups may be more successful in allowing hard to reach populations to participate—such as those in remote duty stations—because they are able to participate from any location using a compatible computer system that is connected to...
the Internet (or an Intranet). For the military this is important due to the fact that our personnel are often mobile and difficult to reach when they are not at their home base.

**BRIEF LITERATURE REVIEW**

**Social Presence**

One of the theoretical approaches that may be helpful in understanding the differences between traditional face-to-face and these newer methods of communication—especially as it relates to the impact of the reduction in non-verbal and social cues introduced by online focus groups—is social presence theory. Social presence is defined as the degree to which a particular communication medium is successful in transmitting both information and social cues present in typical face-to-face interaction between people (Short et al., 1976 as reported in Rice, 1993). The theory encompasses both the content information that is transmitted as well as the impression that there is a ‘real’ person on the other end of the communication. Social presence can be conveyed through written and verbal conversations as well as through nonverbal social and symbolic cues. Media types and interpersonal contact differ in the ability to transmit communication cues, and therefore the ability to convey equal amounts or types of information. A principle underlying the theory is that the more a medium is able to convey social presence the more effective and satisfying the experience will be for participants (Rice 1993; Tu, 2000).

**Level of Participation**

Previous reviews of the literature by McLeod (1992) and Bordia (1997) indicate that reduced social presence may actually enhance the responses of many participants when using online response formats. Both authors cite studies in which participants using online modes of communication provided a higher number of comments (or words used) in the group discussion when compared to those in traditional face-to-face conditions. They suggest this finding is due to the fact that online group discussions allow for simultaneous participation, by letting respondents interject their comments at anytime in the discussion without waiting for others to finish. Previous studies have also indicated that online focus groups tend to yield more unique comments than traditional face-to-face methods. For instance, a study by Parent (2000) found that groups using online discussion formats produced 55 percent more unique ideas than the face-to-face discussion groups. However, it must be pointed out that this study utilized an asynchronous discussion board format and not an interactive synchronous chat room. In addition, Bordia (1997) reported that online group discussions exhibit less evidence of social conformity, which is attributed to the lack social status cues being communicated to respondents.

**Participant Satisfaction**

Findings from several studies suggest that people generally enjoy participating in online group discussions. For instance, Sweeney, Soutat, Hausknecht, Dallin, and Johnson (1997), found that respondents gave higher satisfaction ratings and were more likely to say that they would participate again in a focus group when it was done online rather than face-to-face. However, these authors point out that higher satisfaction ratings may be due in part to the novelty of participating in a new type of group discussion, and not necessarily for reasons such as enhanced quality or ease of communication. For example, Olaniran (1996) found that the ease of use was lower in online group discussions when compared with traditional face-to-face methods; however,
differences in satisfaction ratings between the two groups approached but did not reach statistical significance. Olaniran (1996) concluded that the ease of use of a focus group method was the single greatest contributor to satisfaction with the method. Overall, previous research has found that the majority of participants enjoy online focus groups—even though they appear to be more difficult to participate in (Gunawardena and Zittle, 1997; Tu, 2000).

Problems With Previous Literature

Although there is a great deal of interest, very little research has been published on either the utility or success of conducting online focus groups. Most studies comparing online and face-to-face methods of communication have centered only on brainstorming or other idea generation activities rather than on the typical dynamic group discussions associated with true focus groups. Morgan (1998) identifies the fundamental aspects of a focus group as being research-based, focused on a topic of interest, and involving dynamic group discussion. The majority of published studies have either not been true focus groups (Parent et al., 2000; Sweeney et al., 1997; Olaniran, 1994) or have not directly compared online and traditional face-to-face group discussions (Ahern, 1993; Valacich, et al., 1992). A small number of published studies were found that presented information about actual computer-mediated focus group discussions (Schneider et al., 2002; Tse, 1999; Walston & Lissitz, 2000). While these studies reported results of true online focus groups, only one conducted statistical analyses comparing the results of the face-to-face and online focus group conditions.

OBJECTIVES AND HYPOTHESES

The present study evaluated the impact of mode of participation on perceived social presence and a variety of outcome measures of focus group success. The design included two online methods of discussion compared with the traditional face-to-face mode of participation. The three participation modes were designed to represent distinctly different levels of social presence ranging from high to low (as described below in Procedures). This paper represents only a portion of the entire study conducted—focusing on the results as related to the level of social presence, participation and satisfaction. The hypotheses tested in this paper include:

Hypothesis 1: Compared with the face-to-face condition, participants in both online conditions will demonstrate higher levels of participation.

Hypothesis 2: Compared with the face-to-face condition, participants in both online conditions will have lower levels of social presence.

Hypothesis 3: Participant satisfaction with the overall focus group experience will increase as a function of the level of social presence (as defined by mode of participation).

METHODS

Participants

The present study used a convenience sample of junior enlisted Sailors (E-1 to E-3) from the Naval Training Center Great Lakes. Participants were recruited from the current student inventory at the Combat Systems and Engineering training schools. Recruitment was done by the Chief Petty Officer of each of the schools who asked for volunteers to participate in the focus group, and then assigned them to a day and time on
a schedule provided. Of the 162 Sailors recruited a total of 156 Sailors participated in the study. Participants ranged in age from 18 to 37 (Mean=21.4). Eighty-one percent of the participants were male and over 90 percent had been in the Navy for less than 1 year. No compensation was given for participation in the study.

**Procedures**

Sailors, in groups of 6, were randomly assigned to participate in one of three conditions representing different participation modes and varying levels of social interaction: face-to-face, computer-mediated, and Internet-based. Focus group conditions were balanced for time of day and day of week. The traditional face-to-face condition represented the highest level of social presence due to the fact that participants had access to both verbal and non-verbal communication from all other participants. The computer-mediated condition represented a medium level of social presence due to the fact that other participants were co-located in the same room. While participants were required to communicate with others by use of the computer chat room, they were able to see and hear others while typing their responses into the computer. The simulated Internet condition represented the lowest level of social presence because barriers separated all participants so that they were unable to see the others. In addition, to block both environmental noise (e.g., typing, verbal comments of others, ambient noise, etc.), participants were provided with noise-reducing headphones. Both online conditions (i.e., computer-mediated and simulated Internet) utilized synchronous chat software that displayed comments of each participant in near-real time (1 second latency). Each comment was clearly marked indicating which participant had made the comment to the on-going discussion.

The primary elements in common with all of the conditions were that they focused on the same sequence of discussion questions, were limited to 1 hour of discussion, and were led by trained moderators. Two experienced moderators conducted the focus groups, leading half in each condition to mitigate the impact of moderator style on participation. All groups were conducted in a centrally located classroom at the Naval Training Station Great Lakes. Nine groups were conducted in each of the three conditions resulting in a total of 27 groups.

**Measures**

Participants in all of the focus groups were asked a series of questions regarding recreational activities such as: “what on-base activities do you use?” and “what are some of your favorite places to go off base?” The face-to-face focus groups were audio taped and later transcribed. The computer-mediated and simulated Internet focus groups had the entire discussion saved on the chat room host server. Content analysis based on the transcribed discussions was coded for the unique and relevant comments and the frequency of participation. Four independent raters completed evaluation of each of the focus group transcripts. A total of 7 pieces of information were recorded for each participant including: number of participation attempts; number of relevant comments; number of irrelevant comments; total length of comments; length of relevant comments; number of disagreements; total number of comments.

Following the focus group discussion, participants were asked to complete a short questionnaire that collected a few demographics, as well as asking about their experience and comfort level with using computers for communication. In addition, participants were asked to complete two scales
that measured their perceptions of social presence and their ratings of satisfaction with the overall focus group experience. The social presence scale was made up of 13 items derived from the work of Rice (1993) and Tu (2000) that represent different aspects of social presence (i.e., interactivity, immediacy, interpersonal warmth, sociability, interpersonal connection, etc.). Participant satisfaction was measured by a scale of 7 items that asked participants how much they enjoyed participating, the ease of participating in the discussion, and their willingness to take part in future research activities similar to the present study. Both the social presence and participant satisfaction scales utilized 5-point Likert-type agree/disagree scales. The responses to these scales were averaged to maintain the original scaling.

RESULTS

To test the first hypothesis about level of participation, the researchers used the scores generated by the coding of the transcripts. Two of the scores generated in this process included measures of level of participation: number of participation attempts and total length of comments. A one-way ANOVA comparing the first level of participation measure—number of participation attempts—showed no significant differences, $F (2, 153) = 0.45, p > .05$. However, a one-way ANOVA comparing the second level of participation measure—total length of comments—was significant, $F (2, 153) = 0.45, p > .05$. Post-hoc mean comparisons using the Tukey test revealed that participants in the face-to-face condition used a significantly higher number of words to express themselves than either the computer-mediated (difference = 609, $p < .05$) or simulated Internet conditions (difference = 641, $p < .05$).

The second hypothesis suggested that the level of perceived social presence by the participants would be lower in the online modes of participation in the study when compared with face-to-face. A one-way ANOVA comparing the level of perceived social presence across the three conditions showed no difference, $F (2, 152) = 0.56, p > .05$.

The final hypothesis was that participant satisfaction would increase as a function of the level of social presence (as defined by mode). A one-way ANOVA comparing the level of participant satisfaction across modes showed no difference, $F (2, 152) = 0.04, p > .05$. The results indicated that there were no differences in participant satisfaction as a function of mode of participation.

<table>
<thead>
<tr>
<th>Table 1. Average Scores for Measures of Participation, Social Presence and Satisfaction by Mode of Participation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Traditional face-to-face</strong></td>
</tr>
<tr>
<td>Number of participation attempts</td>
</tr>
<tr>
<td>Total length of comments</td>
</tr>
<tr>
<td>Perceived social presence</td>
</tr>
<tr>
<td>Participant satisfaction</td>
</tr>
</tbody>
</table>

DISCUSSION

Present study sought to compare traditional face-to-face with two online methods of conducting focus group to evaluate the differing levels of social presence and the
overall success of the approach. In contrast to what was expected, social presence did not appear to clearly distinguish between each of the different modes of participation. While there were no differences among on the number of communications attempts, it is clear that the average number of words used in participation was higher in the face-to-face condition. Even though the three participation modes were designed to be different on a number of features, participant social presence ratings did not differ as predicted. Regardless of the mode in which they participated, study participants appeared to enjoy the focus group experience as evidenced by their high ratings of satisfaction. Overall, the results indicate that conducting online focus groups can be done in a successful manner and achieve similar levels of social presence as typically found in face-to-face approaches.

On a practical level, the moderators for the present study often found it more difficult to lead online focus group discussions. Even though participants were asked to communicate by only using the keyboard on their computer terminal, laughing and comments were made by participants during the course of the focus groups. Also side conversations between the participants frequently took place while others were conversing with the moderator. It is possible that the age of the Sailors and their previous experience in chat room discussions played a role in participant’s tendency to be off topic. In addition, the novelty of participating in an exercise that took these junior enlisted Sailors out of the classroom and allowed them to use computers may have negatively impacted the ability of our participant measures to adequately differentiate between the various modes of participation.

Future studies should seek samples that are more likely to represent the people who are likely to be used in this type of data collection in the future. For instance, more senior personnel are more likely to be involved in organizational climate or change research in the military. This older population may respond differently than those who were selected for inclusion in the present study for a wide variety of reasons. Another potential area of research on this topic area includes ‘real-life’ testing of Internet-based focus group discussions.

REFERENCES


CLUSTER SAMPLING VERSUS STRATIFIED RANDOM SAMPLING OF U.S. NAVY PERSONNEL: A COMPARISON

Carol E. Newell & Paul Rosenfeld
Navy Personnel Research, Studies, and Technology Department (PERS-14)
5720 Integrity Drive, Millington, TN, USA 38055-1400
carol.newell@persnet.navy.mil

ABSTRACT
Over the past ten years, response rates on large-scale Navy surveys have steadily declined. Response rates on the Navy Personnel Survey (NPS), dropped from 52% in 1990 to 33% in 2000. This has led to the Navy seeking alternatives to the traditional mailout survey. The purpose of this study is to determine the feasibility of utilizing cluster sampling to develop a statistically representative sampling strategy of Navy respondents for data collection at selected U.S. Navy installations. This study will also statistically compare the demographics of a cluster sample and a stratified random sample to determine if differences exist between the two. The study results and future directions for cluster sampling in the Navy will also be discussed.

INTRODUCTION
U.S. Navy leadership frequently requires information on personnel members’ attitudes and opinions and this data has typically been assessed through traditional mailout surveys (Newell, Rosenfeld, & Culbertson, 1995; Rosenfeld, Newell, & Le, 1998). Over the past ten years, these large scale Navy surveys have suffered a steady decline in response rates (Olmsted, 2001; Rosenfeld & Newell, 2000). The Navy Equal Opportunity/Sexual Harassment (NEOSH) survey, which is mailed biennially to a Navy-wide sample, received a 60% response rate in its first year of administration, 1989. By 1999, the rate fell to 30% (Rosenfeld & Newell, 2000). Decreasing response rates have been noted with concern by Navy leadership, who have requested practical solutions to the problem.

While mail survey response rates have steadily declined, the costs associated with these surveys have increased due to rising postage rates and the larger sample sizes and follow-up mailings needed to ensure reliable samples. These problems associated with mail surveys require that alternative data collection techniques be explored. It has been suggested that these problems could be addressed by administering surveys at the installation or command level. However, there has been no prior determination in a Navy setting that command-level administration is scientifically sound and can be reliably generalized to the Navy as a whole. This study seeks to develop a statistically representative, scientifically valid sampling strategy of Navy respondents through data collection at selected Navy installations. The use of a scientifically valid sample will allow the results to be generalized to the larger Navy population. This study will adapt cluster sampling techniques to ensure that representative samples of the Navy are generated at selected installations.

The opinions expressed are those of the authors. They are not official and do not represent the views of the U.S. Navy Department. The authors appreciate the assistance of Jill Dever.
Cluster sampling is a probability sampling technique where a defined population or geographic unit is split into groups or clusters, a random sample of these clusters are selected from the units, and all observations in the random sample are included in data collection (Fowler, 1993). There are three types of cluster sampling, one-stage, two-stage, and multi-stage. In one-stage sampling all observations in the random sample are selected, while in two-stage cluster sampling once clusters are identified a random sample of the clusters are selected for data collection. Multi-stage cluster sampling differs in that subjects are stratified on a characteristic (e.g., gender, race, job position) that is relevant to a study, and a random sample is subsequently taken from each stratum (Sullivan, Bordiga, & Carter, 1988). Cluster sampling is the recommended sampling technique when the population of interest is widely dispersed geographically and/or data collection will take place on-site (Fowler, 1993; Morgan & Harmon, 1999; Sullivan et al., 1988).

Since a large percentage of Navy personnel are usually deployed at sea and therefore inaccessible for on-site data collection, Navy researchers have typically employed mailout surveys that utilize simple or stratified random sampling so that they can accurately sample the entire Navy population. A few non-Navy military studies have used cluster samples to investigate issues related to youth and veterans perceptions of the military and military service (Bray, Ostrove, Immerman, McCalla, & Guess, 1986; Boyle, Brounstein, Knain, 1983). More recently, a U.S. Department of Defense (DoD) study used cluster sampling to determine the prevalence of substance abuse and other health-related behaviors in the military (Bray, Fairbank, & Marsden, 1999). In this continuing study surveys are administered every three years both on-site at military installations and through the mail for those in remote areas. While the studies described above used cluster sampling, no military study to date has systematically compared a cluster sample to a stratified random sample in a military setting.

The present study seeks to determine the feasibility of cluster sampling to develop a representative sample of Navy personnel for data collection at selected Navy installations. This study will statistically compare the demographics of a cluster to that of a stratified random sample. While cluster sampling has successfully been used in civilian survey settings, few military or civilian studies to date have compared results collected through random samples to cluster samples.

The present study uses the Chief of Naval Personnel (CNP) Quick Poll to compare cluster with stratified random sampling. The CNP Quick Poll was developed to rapidly provide Navy leadership with statistically valid Sailor attitude and opinion data through brief, focused, web-based surveys. For the first poll, two samples of Navy personnel were extracted using stratified random sampling and cluster sampling techniques. The demographic comparisons between the two samples will be presented and a discussion of the feasibility of this technique for other Navy-wide surveys will be discussed.

METHOD

Two samples of Navy personnel were extracted from Navy electronic personnel files using cluster sampling and stratified random sampling techniques. For both samples, the stratification variables were military status (officer/enlisted) and location (sea/shore). The stratified sample included 3,933 personnel members while the sample
size for the cluster sample was 5,660 (see Table 1). The sample sizes were determined using the Sample Planning Tool (Mason, Wheeless, George, Dever, Reimer, & Elig, 1995). This software uses a Karush-Kuhn-Tucker-based numerical algorithm for computing the sample size and allocation required to satisfy a set of precision constraints. This algorithm computes an allocation that minimizes a specified cost model while meeting or exceeding the required precision constraints (Mason et al., 1995). The tool is commonly used to determine sample sizes for large scale DoD and Navy personnel surveys.

| Table 1. Sample Sizes for Cluster and Stratified Random Samples |
|------------------|------------------|
| Stratification   | Cluster Sample   | Stratified Sample |
| Group            |                  |                  |
| Shore Enlisted   | 1,452            | 1,126            |
| Shore Officers   | 1,518            | 997              |
| Sea Enlisted     | 1,900            | 1,347            |
| Sea Officers     | 790              | 463              |
| Total            | 5,660            | 3,933            |

For the cluster sample, a stratified two-stage cluster design was utilized. The clusters were developed by sorting the Navy population by zip code and Unit Identification Code (UIC). A total of 199 clusters (88 shore and 111 sea) were created; from this 11 shore clusters and 10 sea clusters were randomly selected for inclusion in the present study, as determined by the Sample Planning Tool. 5,660 personnel members were then randomly selected from these clusters. For the stratified random sample, 3,933 personnel members were randomly selected.

RESULTS

The demographics of the cluster and stratified random samples are presented in Table 2. As displayed, few differences were found between the two samples. Virtually identical percentages were selected for location status and gender. Chi square tests were conducted and showed that significantly more of the lower paygrade enlisted (14.2%) were selected in the stratified random sample than in the cluster sample (11.4%), while a larger percentage of senior officers was found in the cluster sample (18.4%) than in the
stratified random sample (15.8%). In terms of race, the stratified sample had a significantly higher percentage of Other (17.6%) than the cluster sample (14.6%). The cluster sample had a significantly higher percentage of Caucasians (70.0%) than the stratified random sample (66.1%).

An additional consideration is the number of contacts that would need to be made to successfully deploy the polls. In order to quickly deploy the polls the Navy Messaging System will be utilized for this study. The system allows for the messages to be sent to each command quicker than traditional mailing techniques. One drawback is that each UIC in the sample would need to be entered in order to send the messages. The stratified random sample contained a total 1,560 unique UICs. In comparison, the cluster sample contained 308 UICs.

<table>
<thead>
<tr>
<th>Table 2. Demographics of Cluster and Stratified Random Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Location Status (Sea/Shore)</strong></td>
</tr>
<tr>
<td>Shore</td>
</tr>
<tr>
<td>Sea</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td><strong>Paygrade</strong></td>
</tr>
<tr>
<td>E1-E3</td>
</tr>
<tr>
<td>E4-E6</td>
</tr>
<tr>
<td>E7-E9</td>
</tr>
<tr>
<td>W2-W4</td>
</tr>
<tr>
<td>O1-O3</td>
</tr>
<tr>
<td>O4-O6</td>
</tr>
<tr>
<td><strong>Race</strong></td>
</tr>
<tr>
<td>Caucasian</td>
</tr>
<tr>
<td>African-American</td>
</tr>
<tr>
<td>Other</td>
</tr>
</tbody>
</table>

<sup>a</sup> = significant difference from stratified random sample

**DISCUSSION**

The purpose of the study was to test the feasibility of using the cluster sampling technique to create statistically representative samples of the Navy population. The demographics of a cluster and stratified random sample were compared and few differences were found between the two. On the key demographic dimensions of location, gender, paygrade and race the demographics between the cluster and stratified random sample were very similar. Thus, at the sampling allocation level the argument can be made that samples drawn from cluster designs can approximate the Navy
population at least as well as those drawn using traditional stratified random samples. For this particular application, an additional benefit of the cluster samples was that significantly fewer UICs were included than in the stratified sample, which would aid in notifying personnel members through the Navy Messaging System.

On a statistical level, one potential drawback of cluster sampling is that these samples may be more homogeneous than simple or stratified random samples (Ray, 1983). This means that members of clusters (e.g., personnel on a Navy ship) may be more similar to one another than members of random or stratified samples. While raised as an issue, few studies have empirically tested this. Ray (1983) examined the homogeneity issue in his study on attitudes towards the environment in Australia. In this study the results of data collected via random and cluster samples were compared. Ray found that the cluster sample had a higher internal consistency than the random sample and was therefore more homogeneous. However, since the means on the survey items between the two samples were not significantly different, the author concluded that while the cluster sample was more homogeneous than the random sample, this homogeneity did not bias the survey results. A follow-up study should be conducted to determine the degree of homogeneity present in military cluster samples. Given the diverse populations on most military installations (i.e., officers, enlisted, civilians, etc.) this may not be a large threat to the validity of the survey findings.

In sum, this initial test indicates that cluster samples are a viable alternative to stratified random samples. Future studies using actual operational Navy personnel surveys are required to provide a more rigorous test of the comparability of data collected through these sampling techniques in Navy settings.

REFERENCES


A Cross-cultural Analysis of Distance Learning in the U.S. and India

Richard B. Modjeski, Ph.D.
Florida Institute of Technology
Paxtent River, MD

Raymond O. Waldkoetter, Ed.D.
Educational & Selection Programs
Greenwood, IN

Abstract

Two courses in Computer Science (Artificial Intelligence) were taught using the same textbook and courseware. Graduate students from the Florida Institute of Technology (FIT) [N = 7] and Executive MBA students at the Xavier Institute of Management at Bhubaneswar, India (XIMb) [N = 20] took the courses by Distance Learning over the Internet. Post course Quantitative and Qualitative analyses were conducted on students’ completions [FIT (N=6) and XIMb (N= 15)], homework submissions. The results appeared to support more directed courseware support for students at the beginning weeks of the courses.

Introduction

Knowledge has doubled every 17 years since the time of Newton in the 16th century while digital computers have doubled in processor speed every 18 months according to Moore’s Law from 1946 to the present (Denning & Metcalfe, 1998). This disconnect is exacerbated by the difference between the analog computer which is similar to human brain and the digital computer which approaches continuous change by systems of differential equations. Van Doren (1991) suggested the difference between analog and digital is similar to the distinction between measuring and counting.

Representation of these processes involves deterministic linear measures for digital devices and approximate sigmoid curves for analog devices. Alan Turing (1936, 1950) developed mathematical notions for digital machines to produce outcomes indistinguishable from human analog computing (cognition). Artificial Intelligence (AI) uses these notions in a wide range of applications including education.


The University of Wisconsin Extension (2002) suggests that distance learning may be simply defined as “instructional delivery that does not constrain the student to be
physically present in the same location as the instructor.” This simple definition, however, does not reflect the complex processes and array of technologies mentioned above that are involved in distance learning.

Method

Two distance learning classes in Computer Science (AI) were conducted during the summer term of 2001. A class at the Florida Institute of Technology (FIT) was compared with the same class using the same books at the Xavier Institute of Management in Bhubaneswar, India. Both populations were graduate students with engineering or computer science undergraduate backgrounds.

Graduate students from the Florida Institute of Technology (FIT) [N = 7] and Executive MBA students at the Xavier Institute of Management at Bhubaneswar, India (XIMb) [N = 20] took the course by Distance Learning over the Internet. Course completions were six for FIT and 15 for XIMb students. The instructor used computer hardware, software, and internet facilities to teach both courses from his office in Maryland. The FIT course started 30 April 2001 and ended on 10 August 2001. The XIMb course started 15 August 2001 and ended on 25 February 2002.

The Russell and Novig (1995) textbook was used for both courses. The syllabi for both courses had grading allocated to (Exercises (EXs) and Tutorials 25%; Midterm Examination 25%; Final Examination 25%, and Paper 25%). The following textbook units with exercises were covered:

1. Philosophical Foundations (EX1)
2. Introduction (EX2)
3. Intelligent Agents (EX3)
4. Solving Problems by Searching (EX4)
5. Agents that Reason Logically (EX5)
6. First-Order Logic (EX6)
7. Building a Knowledge Base (EX7)
8. Planning (EX8)
9. Uncertainty (EX9)
11. Making Simple Decisions (EX11)
12. Learning in Neural and Belief Networks (EX12)
13. Robotics (EX13)

Results

Measures of assignments and course completion were analyzed with non-parametric tests using the Statistical Package for Social Scientists (SPSS, 1996). Internet calculators were used to perform Chi Square and Fisher exact probability tests.
A Chi Square test between FIT and XIMb showed no significant differences in course and assignment completion rates (P > .01). Overall completion rates were similar for the two groups. It may be noted, however, that the XIMb class did not complete EX7 (Building a Knowledge Base) and EX9 (Uncertainty) due to scheduling conflicts associated with graduation. The removal of EX7 and EX9 did not appear to change the trend of the analysis.

Course submissions with the exception of the final examination were unscheduled. A Fisher exact probability test was performed on submission times of the midterm examinations. A 2 x 2 table analyzed the time the midterm examination was submitted. The midpoint of each course was used as a measure of speed of submission of the midterm examination. Midterm examinations submitted before and after the course midpoint were analyzed. FIT’s 15 week midterm examination submission was compared with XIMb’s midterm examination submission. XIMb’s 29 week course midpoint was 8Dec01. FIT’s 15 week course midpoint was 16Jul01. FIT’s midterm examinations submitted before the course midpoint = 2 and after the course midpoint = 4. XIMb had three midterm examinations submitted before the course midpoint and 12 midterm examinations submitted after the course midpoint. The Fisher exact statistical test of the results of midterm examination submissions for both schools was significant (p < .05).

Discussion

![Figure 1: Machine vs. Human Performance as a Function of Time](image)
Figure one illustrates rough plot of student performance (Midterm Examination Turn-In) over time follows the sigmoid (S-shaped) learning curve. The expected performance of a machine hardware/software system follows a (straight line) linear curve over time.

The areas intersected below and above the linear curve may be areas of intensive Computer – Human Interaction (CHI). The lower area starting at the first positive inflection point of the sigmoid curve and ending at the point of intersection with the linear curve may provide a useful starting point for research that allows both computer-based instruction and distance learning over the internet to enhance human performance.

The most dropouts from the courses occurred early in the semester before the students may have had time to organize the material. Computer scheduling of lessons and assignments may provide greater understanding by the student and lessen the motivation to drop the course. The use of Instructional Technology to design courseware that translates the goal of keeping the student on task and on track may enhance curriculum delivered by distance learning.

The instructional design needs to go beyond repetition and reinforcement (Ausubel 1978; Gagne, 1977) to be meaningful to the student. Basic research in the analysis of transfer of defining attributes of concepts (Bourne, 1971) may provide a methodology for building better distance learning courseware.

1. Author Note: The authors thank Professor William B. Michael Ph.D. of the University of Southern California for his suggestions on statistical analysis.

References


Fisher Exact Test: http://www.physics.csbsju.edu/stats/fisher.form.html


Web Chi Square Calculator: http://epsilon3.georgetown.edu/~cball/cgi-bin/webchi/web_chicgi

Commander Marten Meijer Ph.D. Giel E. Algra MSC, Jos M.P. Weerts MD?

Summary
During the past decades, more than 80,000 military personnel of the Netherlands Armed Forces participated in international peacekeeping operations. In caring for them in a professional way, many lessons have been learnt and personnel policy has been strongly improved. In this paper this personnel policy is being examined, especially in the research that is being done among military personnel after deployment. In particular the non-response in survey-research among them will be discussed as well as self-harm and violence after deployment. Recommendations are stated for further improvements of policy on mental health care of military personnel and veterans.

Introduction
From 1979 on, the Netherlands Armed Forces participated in the United Nations Interim Forces in Lebanon (UNIFIL) for a period of six years. This UNIFIL deployment was the start of a new era, after the departure of Netherlands Armed Force from New Guinee in 1962. In between, Netherlands Armed Forces had only been training for a large scale conflict with the Warsaw Pact Forces.

In this paper we will review the policy and facts on care of military personnel after deployments from UNIFIL on. We will focus on the subjects of response in mental health surveys and reunions. In combination with the phenomenon of overt violence, self-harm and suicide after deployment, we will draw conclusions about the lack of information on military mental health after deployment. We will end our paper with recommendations on countering non response in surveys and reunions, collecting data on military mental health after deployment and care for veterans in general.

In this paper veterans are defined as military personnel that has been deployed in active service and has left the active service. Figure 1 shows the difference in definition between veterans and non-veterans.

? The Centre for Research and Expertise of the Veterans Institute, P.O. Box 125 3940 AC Doorn The Netherlands.
Telephone ++ 31-343-47 42 57
Fax ++ 31-343-47 42 54
E-mail: koc@veteraneninstituut.nl
www.vi-koc.nl
Figure 1. The Netherlands definition of veterans: military personnel deployed and out of active duty.

From figure 1 it appears that deployment is the distinct feature of veterans. Deployments enclose warfare, service in warlike circumstances and participation in peace-operations. This raises the question about the differences between waging war and participating in peacekeeping or peace-enforcing operations, especially in the context of the psychological aspects of war and peace-operations.

At first glimpse, war has a far more violent character. Often the goals in war are more distinct and more clearly national interests are at stake. From a psychological point of view, these characteristics might have a more profound impact on human behaviour. However, violence during war is most of the time well planned and in accordance with the goals to be reached. This might reduce the psychological impact of war both in attack and defence in comparison to peace-operations, in which violence is always unforeseen and takes place suddenly and unexpected. The violent stages in a war also present an outlet to the psychological arousal of soldiers, while in peace operations military personnel is aroused without the opportunity of expressing this arousal in violent behaviour.

Secondly, the observation that goals in a war are more clearly defined than in peace-operations can also explain a reduction of the psychological burden of waging war. Especially when land or sea areas have been conquered or enemy positions have been destroyed, there is an overt reward of military behaviour. In peace-operations there is no
such reward. Furthermore, when rules of engagement specifically forbid to be violent, there is no reward of military behaviour at all, with a far greater chance of frustration and all kinds of psychological injuries.

In the third place, in waging war often more clear national interests are at stake than in peace-operations, fostering commitment among national military personnel and probably making them less vulnerable to psychological injuries. In sustaining the war and aslo in recuperation after a war these national interests also will contribute to societal esteem for veterans who have been in charge of these national interests.

The three remarks as discussed before, do not lead to the conclusion that the psychological impact of waging war is the same as the psychological impact of participating in a peace-operation. Many authors already stated that peace-operations have their own specific stressors, like vulnerability and helplessness (Schoeman 1996, Extra 1998). However, we do conclude that participating in peace-operations has its psychological burden as well, so veterans of these operations deserve care and attention as well as war-veterans. Experiences both from war and peace-operations affect the lives of veterans in the post-deployment period, long after the actual departure from the theatre. Last but not least, we note that from the perspective of psychosocial care it does not make sense to weigh the burden of experiences or loss, but that all of them need their own kind of attention.

Developments in the population of veterans

After the Second World War the Netherlands government had to reduce the armed forces with large numbers of personnel. That became the offspring of large amounts of veterans.

The deployments of Netherlands armed forces between 1947 and 1962 in the East Indies, Korea and the former Netherlands New Guinee enlarged this legion of veterans again with large numbers. The veterans from this era are called ‘old veterans’.

The veterans from UNIFIL and latter operations are called ‘young veterans’. However, a New Guinee veteran who has been deployed at the age of 18 in 1962 is called an old veteran but can be younger than a UNIFIL veteran who has been deployed in 1979 at the age of 48. More important, the genuine contact with an old veteran of 86 can even be impaired by calling him old. Figure 2 shows the the estimated numbers of old and young veterans from 1990 to 2004.
From figure 2 it appears that the number of old veterans is decreasing and the number of young veterans is increasing. It is expected that in 2010 they will reach a break-even point. These developments imply a change of focus from primarily old veterans to young veterans. This shift in focus does not imply that the old veterans are not being cared for, respected or appreciated anymore. For their bravery, courage and their successful efforts to survive and cope with their experiences in the rest of their lives, they deserve all the respect they can get. However, in this paper we will focus on young veterans.

**Policy on military mental health care after deployment**

From approximately 1980 on, the Department of Defence (DOD) in the Netherlands has been trying to increase the mental health care for military personnel during deployments and in the pre- and post-deployment period. Already during UNIFIL social workers have been deployed with the deployed unit and in the post-deployment period mental health was cared for by military psychologists, both in therapy and research. From the UNIFIL timeframe it appeared that a minority of the deployed personnel suffered from Post Traumatic Stress Disorder (PTSD), sometimes even followed by a compassionate leave of the deployment. Especially Netherlands and Norwegian military personnel appeared to develop psychiatric complaints during UNIFIL (Weisaeth, 1994). In the years to come, the DoD increased its activities in mental health care, including a pre-deployment briefing on the effects of stress, the presence of mental health workers during deployment and debriefing, homecoming interviews and mental health surveys in the post-deployment period. The DoD also offers facilities for reunions of the deployed unit, with a reunion every five years of deployed marine corps units. Figure 3 shows these activities of military mental health care during preparation, deployment and return.
Figure 3. Activities of military mental health care during the stages of pre-deployment, deployment and post-deployment.

From figure 3 it appears that nowadays a lot of activities of military mental health care are common practice. Home-coming interviews, carried out by social workers 2 to 3 months after deployment, are still not that common. As a matter of fact, all those activities emerged against all odds, after a long struggle to overcome the military masculinity culture, in which mental health care is still a Fremdkörper. In the remaining part of this paper we will reveal some facts about these activities, observing form the point of view of the Veterans Institute in the Netherlands. In figure 4 the partner-organisations contributing this institute are shown.
From figure 4 it appears that five partner-organisations are contributing to the Veterans Institute. Especially the participation of organisations like the Veterans Union and the Foundation of War and Service Victims creates the possibility to supervise and criticise the DoD whenever that is needed. This kind of co-operation between organisations with apparently different interests is quite common in the Netherlands and has been called the ‘polder-model’. We have been using this model for collecting facts about military mental health care, coming from DoD as well as veterans organisations. We will focus on response in surveys on mental health care, reunions, the quality of mental health care and the self-harm among veterans.

Response in surveys and reunions and the quality of mental health care.
The DoD sends a mental health survey to the deployed military personnel 6 to 9 months after the end of the deployment. The survey consists of more than 250 items and focuses on traumatic events both in deployment and private life. The survey has been sent to 4 marine corps units of approximately 200 members each, that have been deployed in Bosnia and Haiti in 1995 and 1996. A reunion for these units has been held in 2001. Both in survey and reunion the overall response was 40%. However, the analysis of this response revealed remarkable differences between the response in the survey and the response in the reunion (Meijer, Wokke and Weerts 2002, p.13). Figure 5 shows the response on the mental health survey and the reunion, divided into officers, non-commissioned officers and marines and the presence of these categories in the strength of the deployed units.

Figure 5. Percentages of officers, NCO’s and soldiers in unit strength, reunions and surveys among Marine Corps Units, deployed in Bosnia and Haiti in 1995 en 1996.

From figure 5 it appears that there is an over-representation of NCO’s and an under-representation of soldiers in the survey. The response in reunions is more similar to the unit strength. However, policy on mental health care is only based on the surveys, which appeared to be biased by rank. It appeared that soldiers differ form NCO’s in deployment.
experiences and in need for mental health care. Therefore, it is needed that mental health care policy and the evaluation of this policy is not solely based upon mental health surveys.

The Social Work Branch of the Dod (Meijer, 2001) and Veterans Institute (Meijer, Wokke and Weerts, 2002) have studied the quality of social work for personnel of the armed forces, the marine corps and veterans. In figure 6 this quality is rated on a scale from 1, very poor, to 10, excellent. This ratings have also been gathered for meeting colleagues of the deployed unit during a reunion.

![Figure 6. Ratings of quality of contact, effectiveness of help and meeting of colleagues among armed forces personnel, marines and veterans (1=very poor, 10=excellent).](image)

From figure 6 it appears that the quality of contact is rated higher than its effectiveness. The quality of social work for the marine corps can be slightly improved. However, on both aspects of the quality of social work the ratings are quite satisfactory. Meeting colleagues of the deployed unit at a reunion is rated even better. For the policy on mental health care these results indicate that contact between member of the deployed unit should be fostered. In measuring this need for contact, the need for sharing positive experiences of the deployment appears to be the strongest predictor of this need for contact (Mouthaan, Eeuwema and Weerts, 2002).
Self-harm and violence in the post-deployment period. When does the war stop?

In 1975, the United States of America ended their war in Vietnam. More than 3 million American soldiers served in Vietnam mostly for a tour of duty of one year (Shephard, 2001, p. 340). Of these soldiers approximately 57,000 were killed in action, a much larger number was wounded in action and an unknown number is still missing in action. Many stayed physically impaired, due to their injuries. In the years after 1975, a still unknown number committed suicide, or was killed in violent deaths.

In Vietnam, Sergeant Dwight Johnson had won the Medal Of Honour, the United States’ highest decoration for valour, for single-handed knocking out twenty enemy soldiers during a raid on his position. He had then served with distinction for another two years, but on returning home, found difficulty in readjusting to civilian life. He became convinced that the Army exploited all black soldiers and made no effort to help them afterwards: Army psychiatrists did not change this view. His frustration grew until he decided to deploy in his rundown Chicago neighbourhood skills he had shown in Vietnam. He was robbing a liquor store when he was killed...(Shephard, 2001, p.357).

In the case of Sergeant Johnson, courage in action is accompanied by an overt act of cowardiceness after return in civilian life. Less overt or even strictly hidden are some psychosocial aspects of the after deployment life. Not all behaviours from the deployment period as committing war-crimes, atrocities or the mere survival of the deployment are public knowledge. Hendin and Pollinger-Haas (1991) conclude in their research on suicide among Vietnam veterans that especially feelings of guilt, coming from surviving beloved comrades as well as from killing defenceless people like prisoners of war, elderly people, women and children, are the main reason for suicide. Killing out of fear or rage has the most distinct relation with suicide. Attacking enemy villages by order, in which also many civilians were killed, is less strongly related with guilt and suicide. They also found out that Vietnam veterans are between 11% to 65% more likely to commit suicide than non-veterans. Their conclusion that feelings of guilt are the most powerful predictors of suicide resembles findings from research among veterans of the Second World War. The strong benefit of their study is that feelings of guilt appear to be intermediate between having killed people and suicide: these feelings were not identified as such in former research. Especially in psychotherapy for veterans these feelings of guilt have to be worked through, in spite of shame and hesitation of mentioning them. Until very recent, this shame and the importance of guilt do not receive the attention they deserve (Leskela, Dieperink and Thuras, 2002), and are also likely to be taboo.

Behavior in the battlefield: filters, errors of omission and errors of commission.

From research among veterans it appears that in the battlefield situation, a change of mind occurs. When decisions of life and death are at stake the consciousness gets purified. Every distraction is filtered out and the mere survival of the battlefield demands
full attention and utmost concentration. In such a state of mind, sensory perception is working at a maximum, encoding a lot of information from the battlefield together with a huge amount of sensory information. In meetings with veterans this aspects deserves full attention, once and again. The huge responsibility, coming from the ‘life and death’ decisionmaking, might even decrease the value of everything else in life. In returning from an 18 month deployment in the Far East, a Dutch veteran at the age of 22 concluded: ‘In the rest of my life, there won’t be anything more significant than what I have been through in the last one and half year’. The documentary ‘First Kill’ of Coco Schreiber even shows Vietnam veterans addicted to the power of deciding over life and death (Algra and Meijer, 2002).

An important aspect of this change of mind is the filtering of information. In bombarding Baghdad during the first stage of operation Desert Storm in the Gulf War in 1991, a fighter pilot declared in a CNN interview, right after return on his airbase: Baghdad lit up like a Christmas tree, it looked like the fireworks at the Fourth of July. There is no need for interviewing citizens of Baghdad to conclude that his information has been filtered. On the other hand, this filtering of information is necessary to perform and survive. Therefore, in preparation for war, the mindset of military personnel is being cleared in very simple schemes. The enemy is dehumanised, nicknamed and brought back to distinct features, which contributes to a mindset that allows to kill them in large numbers and with the use all of the human intellect in torture and suppression. In the post-deployment period, careful attention should be paid to these filtering mechanisms, in which step by step the filtered view can be enriched by more accurate information. In meeting veterans and therapy for veterans the errors of omission, coming forth from this filtering, can be worked through by paying attention to all sensations of the battlefield. Also the mere asking of the simple question: ‘Is that all, you want to share with me?’ can reduce the errors of omission.

By debriefing events and operations group-wise, errors of commission might be prevented quite easily too. Especially when there is a dominant culture of masculinity, in which performance and pride are being emphasised, the seduction to remember bravery that never took place can hardly be withstood. In the documentary Crazy from Heddy Honigmann the masculine culture of performance and pride is creatively encountered by asking veterans for their memories of the music they often heard when the were in action. A more complete recollection of the situation and a more extended processing of all kinds of reactions is possible then, which is rewarding for both veterans, their therapists and researchers.

Psychosocial aspects of the postdeployment period: unemployment, divorce, posttraumatic stress and suicide

Kramer et al (1992) included into the subject of suicide also life-threatening behaviours, like motor accidents, shootings alike in box 1 and overdoses of alcohol and drugs. They also included thoughts of death and dying into their research. Figure 7 shows some of their findings about thoughts of death and dying and thoughts of suicide among non-patients, a therapy-group of veterans and an outreach group of veterans.
Figure 7. Thoughts of death and dying and thoughts on ending my life among non-patients, veterans in therapy and an outreach group of veterans (Kramer et al. 1992, p.66).

From figure 7 it appears that veterans in therapy think most of death and dying and suicide. Remarkably, the outreach group has the strongest psychosocial problems, like unemployment and divorces, as shown in figure 8.
Kramer et al. do not offer an explanation for the differences between psychological and psychosocial problems among these groups. Probably the fact that in therapy there is much focus on trauma, events in which death often plays an important role, can explain why the in group in therapy reports more thoughts of death, dying and suicide. The outreach-group can have more severe psychosocial problems, because in many cases partners or colleagues of veterans stimulate veterans to go in therapy. Once these partners have been lost by divorce or loss of employment, veterans will not reach therapy. At the end of the day, both figures demonstrate very clearly that veterans in therapy and the outreach group have severe psychological and psychosocial problems.

Wang et al. (1996) describe the cyclical process of Post Traumatic Stress Disorder (PTSD). This psychiatric disorder follows a traumatic event and consists of intrusion of memories of the event, avoidance of triggers related to the event and a high level of arousal, often leading to overt aggression and hostility, even to relatives and friends. In their model Wang et al. connect PTSD with the stages of grief and mourning as described by Horowitz (1978). These stages are emotional outcry, denial, oscillation between reacting and numbing, acceptance and solution (Horowitz, 1978). In these stages, the oscillation between reacting and numbing seems to be most beneficial for reaching the final stage of solution (Epstein, 1989). Meijer and De Vries (2001) concluded on their
help to veterans that especially very opposite reactions on very different aspects of a situation have to be discovered and worked through. For instance veterans who survived an air-raid in their Mitchell B-25 during the Second World War, have to experience the relief of survival (aspect 1), next to the fear of almost being killed (aspect 2) and the grief of losing so many comrades (aspect 3) whom have been less lucky (Meijer and Weerts, 2002). Figure 9 shows these connections between aspects of events and the reactions of veterans, divided into observations, behaviour and feelings.

![Diagram](image)

**Figure 9. Connections between aspects of events and reactions which consist of observations, feelings and behaviours.**

The model for encountering experiences of veterans in meetings and therapy in figure 9 reveals the fruitful connection of facts at the upper level and feelings at the lower level. Both facts and feelings deserve attention in care and recognition for veterans.

**Long term aspects of the post-deployment period: runs, bunkers, violence and prison.**

Wang et al. (1996) found out that in more long-term life-stages, adaptation to demands of everyday life can be followed by a stage of surviving, in which general functioning is being impaired. In crossing the threshold to decompensation, veterans lose their jobs and divorce by behaviour that varies between utmost excitement (runs) and total isolation (bunker), in which every perspective of the future is being lost. The next and last stage is regrouping or getting lost. Veterans regard this last stage to be very similar to the stage after the battle in a combat zone, in which comrades are being found or lost. In addition to the immense feelings of relief and grief from their past experiences, also the run and bunker behaviours add strong feelings of guilt and shame from their present behaviours.
In that stage, for many veterans death by suicide or violent behaviour seems the only way out. We strongly agree with the authors that this cyclical character needs attention in both therapy and research. The authors also note that medication only offers a temporary solution and that the majority of Vietnam veterans has strong relational problems and up to 70% of them has been divorced (Kulka et al, 1990). Therefore we invite other researchers to do more research on long-term effects of PTSD.

From research of the bureau of Justice Statistics in the United States it appears that veterans are in prison for violent offences more often than non-veterans are in prison (Mumola, 2000). Figure 4 shows the percentages of offences of veterans and non-veterans in State prisons in 1997.

![Figure 10. Percentages of offences of veterans and non-veterans in State prisons in 1997. Source: Bureau of Justice Statistics, January 2000.](image)

From figure 10 it appears that 55% of the veterans in State prisons have been sentenced for violent offences, compared to 46% of non-veterans. About 35% of veterans in State prisons were convicted of homicide or sexual assault, compared to 20% of non-veterans. Among violent State prisoners, the averaged sentence of veterans was 50 months longer than the averaged sentence of non-veterans (Mumola, 2000). The increased arousal among veterans, as part of PTSD, can explain this more violent character of some
veterans. It is also possible that they are more apt to the use of weapons as a solution to conflicts, coming from their war-experiences.
Box 2. Self-harm and violence after deployment in the war on terrorism in Afghanistan.

On June 10 2002, Sergeant First Class Rigoberto Nieves, a member of the Third Special Forces Group who had returned from Afghanistan two days earlier, shot his wife and then himself. He had requested leave from duty in Afghanistan to resolve personal problems. On June 29 2002, the wife of Master Sergeant William Wright of the 96th Civil Affairs Battalion, a Special Forces Unit, was strangled. Wright, who had been back from Afghanistan about a month, was charged with murder. On July 19, Sergeant First Class Brandon Floyd shot and killed his wife and then himself, according to investigators. Floyd was identified as a member of the Delta Force, a crack anti-terrorism unit, whose existence is not officially acknowledged. Washington Post, July 27, 2002, page A03.

Wong et al. (2001) conclude on their research among Canadian veterans of UN-peacekeeping operations that these veterans are not more likely to commit suicide than other Canadians of the same age. In our opinion, they neglect the fact that military personnel is selected upon physical and psychological health in this analysis. Therefore, their research hypothesis should predict less suicide among military personnel than among civilians. Upon this hypothesis their results even might support the conclusion that participation in peacekeeping operations increases the likeliness of suicide. It is a pity that the authors ignore this ‘healthy worker effect’, although they mention it in their study. They do conclude that the likeliness of suicide increases among airforce personnel that have been participating in peacekeeping operations. This personnel is being rotated very frequently and often on an individual rotation-schedule. Further on, there are indications that some of the Canadian military personnel participate in peacekeeping operations as often as possible, because of the rewards and bonuses upon their normal wages, that can be interpreted as risk-taking behaviour. These indications need more careful investigation.

Discussion
From the psychosocial and behavioural aspects of the post-deployment period as described above, it can be concluded that a war never stops, not even after it is over. The experiences are of such impressiveness, that they affect a lot of the life in the post-deployment period. In the Netherlands armed forces, a lot of activities of mental health care have been developed. However, the response on these activities has to be improved, including feedback on these activities to the participants. On the results of research among UNIFIL veterans and Vietnam veterans we conclude that these deployments increase psychosocial and psychological problems, especially among outreach groups of these veterans. Therefore, it is recommended that mental health care is intensified, both during deployments and in the pre- and post-deployment period. However, a vast majority of veterans copes with everyday life excellently. Not all of these experiences are that negative and even a lot of negative experiences are being coped with in an excellent way. People can even grow and mature from their experiences in
war: they have learnt to risk their lives and survive in very difficult and threatening situations. They have been connected to their comrades in a very intimate manner and share a lot of unique experiences. They often benefit from their war experiences and are very successful in everyday life. One of the former Prime Ministers of the Netherlands is a veteran, who has been commanding officer of a submarine during the Second World War.

References


MULTIMETHODOLOGY IN PRACTICE: AN APPLICATION IN THE PORTUGUESE ARMY PERFORMANCE APPRAISAL MODEL.

Mónica A.L.S. Leitão¹
João P.C.F Thomaz²,³
Carlos A. Bana e Costa³,⁴

¹Lusíada University, Rua da Junqueira, Lisbon, Portugal
²Portuguese Military Academy, Paço da Rainha, 1150-244 Lisbon, Portugal
³Centre of Management Studies (CEG-IST), Av. Rovisco Pais, 1049-001, Lisbon, Portugal
⁴London School of Economics, Department of Operational Research, Houghton Street London WC2A 2AE, U.K.

Abstract: Over the last ten years, the Portuguese Army’s officers have been evaluated annually by their commanders, who have used a simple notation model which includes 16 attributes and a specific 5-level descriptive scale for each one. This model has become ineffective for several reasons, mainly because it is no longer able to distinguish between very good and good performers. The case presented searches to enable the Portuguese Army to identify the cases of inadequate performance of the army officers, in a clear and simple manner, in order to introduce efficient corrective actions.

This paper describes the intervention process in which Decision Conferencing and Process Consultation techniques were used to help a working-group in structuring a new competency based performance appraisal model for the Portuguese Army, as well as the construction of a qualitative additive model with the support of the MACBETH approach, to evaluate the identified performance appraisal model.

Keywords: Performance Appraisal, Decision Conferencing, Structuring, Evaluation, MACBETH, Army.

1. Problem Structuring

The decision problem addressed in this paper is the definition of a competency-based performance model for the evaluation of Portuguese Army Officers. This performance appraisal model should:

- Evaluate the performance of the military officers in a simple and clear manner.

Actors and boundaries of analysis

A multidisciplinary Work Group (WG) of eight experts, both military and non-military, was created for the specific purpose of studying the problem, of reviewing the performance evaluation model in use and to elaborate recommendations leading to its improvement. We acted as decision-analysts and facilitators for the Work Group.
The scope of the study was limited to the Portuguese Army Officers evaluation process since NCO’s and soldiers would require a different sort of analysis, based on a different set of indicators.

**Objectives of the study**

The main objectives of the study were: (1) to identify key performance concerns to be taken as evaluation criteria; (2) to develop a transparent and consistent model of performance appraisal for the Portuguese Army Officers and (3) to contribute to the diffusion of Decision Analysis within the Portuguese Army, particularly with regard to Human Resources Policy.

**The Extant Model**

The model in analysis was implemented in 1990 due to the need of the Army regarding the identification of its best performers, taking into account the restructuring and downsizing brought about by the aftermath of the war in Africa (1961-1974).

After ten years of continuous application it would appear that this model no longer serves its initial purpose: a) the evaluation criteria it presents are dubious and conditioned by the individual criterion used by each evaluator; b) sometimes evaluators have made a biased use of this instrument, namely through a tendency to evaluate subordinates in the same manner as they themselves were evaluated, that is, if the evaluator had received a good grade then he would also evaluate his subordinates in that way; c) on the other hand it often happens that the manner in which subordinates were evaluated dictates the evaluation which Commander will in turn have, and finally d) there is a tendency to use only the right side of the evaluation scale making no difference between average (normal) and superior performance.

**Definition of the Problem**

The fundamental starting point was to achieve a *shared understanding* of the actual problem from the actor’s perspective and to develop a *common sense of purpose*. (cf. Phillips, 1984a)

In a Decision Conferencing framework the facilitators started with “pure inquiry” questions (Schein, 1999) in order to sharpen and highlight aspects of the problem presented, and to test the Work Group willingness to consider them in a proper light.

It quickly became clear that the WG intended to define and develop a performance appraisal model that would focus on certain competencies, believed to be the basis for personnel potential development.

“Competencies were defined as general descriptions of the necessary abilities to perform successfully in specified areas. Abilities are the factors that are fundamental for successful on-the-job performance. **Competency profiles** synthesize skills, knowledge, attributes and values, and express performance requirements in behavioural terms” (PSC’s Learning Resource Network).
Thus, competencies are related to the effective performance of a particular job within a specific organizational environment and can be the basis for the Human Resources Policy as a whole, since it affords a clear common understanding of what is needed and wanted in order to insure organizational success.

Overview of the methodology adopted

To guide the work process of analysing and evaluating performance evaluation criteria, we proposed the adoption of a multimethodological approach (cf. Mingers and Gill, 1997). As such we proposed the use of:

1. Multiple Criteria Decision Analysis (MCDA) methodology, which can be described as a package of activities grouped in three main phases of analysis (cf. Bana e Costa, 1999):

   a) Structuring – this phase includes the activities related with problem-structuring (the definition of the problem), model-structuring (definition of the basic elements of the model) and impact assessment (implies the qualitative analysis of the problem);

   b) Evaluation – this phase includes the construction of a qualitative additive value model (cf. von Winterfeldt and Edwards, 1986, Belton, 1999) with the support of the MACBETH approach (cf. Bana e Costa and Vansnick, 1999, 2000), options-evaluation (to evaluate the intrinsic attractiveness or unattractiveness of the options), sensitivity and robustness analysis (to guarantee that the model is a “requisite model” – cf. Phillips, 1982, 1984a);

   c) Elaboration of recommendations – this is not actually a phase in itself, as the elaboration of recommendations is present throughout the decision-aid process.

This helping process evolved in a Decision Conferencing framework, guided by Process Consultation principles.

Decision conferencing (cf. Phillips, 1982, 1984a) is a two or three day meeting to solve relevant issues, which is attended by key players who represent the diversity of perspectives on the question under analysis. Impartial specialists on group process and decision analysis facilitate this process. This methodology implies the use of a “requisite model” (when its form and content are sufficient to solve the issues of concern) created in situ to help provide a structure for the thinking through of the problems at hand.

Its general objectives are:

?? To generate a shared understanding of the issues (not necessarily unanimity);

?? To develop a sense of common purpose (allowing for individual differences of opinion);

?? To agree about a way forward (a commitment to the direction not to individual paths).
Process Consultation “(...) is the creation of a relationship with the Client that permits the client to perceive, understand, and act on the process events that occur in the client’s internal and external environment in order to improve the situation as defined by the client”. (Schein, 1999).

Put together, these activities constitute a multimethodological recursive and learning group-process, which evolved during a series of intensive “face-to-face” facilitated workshops and decision conferences. The initial meeting with the Work Group ended with an outline of the proposed methodology and a discussion of its component activities.

Multimethodology refers to the combining of more than one methodology in whole or in part, or techniques from different methodologies within a particular intervention (cf. Mingers and Gill, 1997). In our case we propose the portioning of methodologies into component parts, being these combined together to construct an ad hoc methodology suitable for our particular problem.

2. Model Structuring

Identifying the key concerns to be taken as evaluation factors

The first workshop devoted to model-structuring comprised several working sessions of a few hours each.

We started by benchmarking a few other military performance appraisal instruments in order to identify common competencies, such as the British, the Brazilian, French, Italian and the Spanish, to determine which evaluation factors were being used for the evaluation of the military personnel.

We then invited each participant individually to write on “post-it” (also known as “Oval Mapping Technique”- Rosenhead and Mingers, 2001) short descriptions of performance evaluation factors thought to be important for the evaluation of military officers independently of rank, cross functionally.

These “post-its” were then stamped on a white board. The following items were identified:
A discussion was then launched as it became clear that many participants had similar concerns related to the problem under analysis. This enabled the facilitators to group these areas of concern into sub-sets of similar or related aspects, by asking participants “why?” and “for what?” each aspect was important. (“diagnostic inquiry” – Schein, 1999).

This process proved to be very helpful as it enabled the participants to identify new aspects and relationships of the problem. Different aspects were re-described, disassembled and linked. By the end of this session the following map was presented.

The next work session started with the analysis and discussion of the map and with the clarification of each of the key concerns identified before. The facilitators stimulated
“dialogue” (Schein, 1999) between participants in order to clarify what each of them meant regarding these key concerns, and why they were perceived as important enough to be mentioned. At this stage participants became aware that in many cases it was not clear to themselves what they exactly meant when using each concept.

One of the most interesting examples was the case of “Commitment to Military Values”, which was considered to be a key orientation of military conduct, but very difficult to circumscribe. It thus became clear to the participants that not all key concerns were simple and clear, making measurement difficult if not impossible.

To incite dialogue between the team members, the facilitator asked one of the members what specifically was bothering him. He felt that not all relevant evaluation factors were being taken into account such as, “courage” or “personal development”. Immediately, other members intervened, presenting more alternatives, such as, “making decisions” and “planning and organisation”.

Thus, it became necessary to go back a few steps in order to reconsider intentions and to revise what had been done. This prompted new discussion and re-evaluation of the steps taken before, especially with regard to the importance of the various items that were identified and their exact meaning. As a result, many key concerns were dropped or re-named, whereas new concerns were put forward. The recursive characteristic of the learning process thus becomes quite visible.

By the end of this session the following competencies were brought forth.
The following session began with the re-analysis of the key concerns (competencies) in order to determine whether a shared understanding of these concepts had been reached. The participants felt that this had been accomplished.

The next step implied the grouping of these competencies into sub-sets of related items (competency clusters). The facilitators intervened in a “doctor-patient” framework (Schein, 1999) presenting their vision of competency profiling and the way in which competency clusters could be easily identified. This type of intervention (“confrontive inquiry”), forced the participants to think about the situation from a new perspective. As a result new “dialogue” was stimulated as a means of reaching “consensus” (which implies that all participants are made to feel that they had a fair chance of influencing the decision) on the identification of the competency clusters and of the competency profile being defined.

The process was quite helpful and enabled the identification, at the end of this session, by means of “consensus” (Schein, 1999), of the following competency clusters:

a) Technical Competencies – refer to using and applying the appropriate technical and operational skills and experience;
b) Interpersonal Competencies – refer to respecting of other people’s abilities, opinions and feelings and to responding appropriately to questions and challenges from others.
c) Personal Competencies – also called self-management competencies, refer to one’s self-control, self-confidence, initiative, etc.;
d) Institutional Competencies – refer to the specific organizational culture.

The following session comprised another recursive moment of re-evaluation of the steps taken, especially in order to clarify possible discomfort with relation to the model that was being defined. This re-evaluation listed some inconsistencies with regard to the inclusion of each individual competency in its respective cluster, as some competencies proved to be germane to more than one competency cluster. The facilitators stimulated “dialogue” between participants to clarify these difficulties, which were easily overcome. As a result the following competency profile was identified:
a) Technical Competencies:
1. Professional Effectiveness;
2. Problem Solving;
3. Decision Making;
4. Planning and organizing.

b) Personal Competencies:
1. Self and subordinate development,
2. Determination;
3. Initiative (proactivity).

c) Interpersonal Competencies:
1. Leadership,
2. Team-work (being a team player);

d) Institutional Competencies:
1. Readiness to serve;
2. Courage;
3. Loyalty;
4. Discipline.

The next step was the definition of the method to be used in the measurement of each competency (construct). Making the constructs (competencies in this case) operational implied defining their meaning and their composing dimensions, so that measurement would become possible. It also made it necessary to define observable behaviour that would indicate the presence of particular constructs.

At this point, the facilitators abiding by the “process consultation model” opted not to intervene. By doing so, the facilitators enabled the participants to evaluate by themselves “what they really wanted” and “how” they wanted it. This was a moment in the helping process that “helped the participants help themselves”, to think through analyse, and solve, the problem. It is essential in a process consultation environment not to substitute oneself for the “client” to whom both the problem and the solution belong.
The participants decided that the best way to measure the identified competencies would be to define statements (direct statements definitions) that would indicate observable behaviour pertaining to each competency.

The first competency considered was “Problem Solving” which refers to the capacity for identifying problems, for collecting relevant data for analysis and for obtaining relevant information in order to solve problems efficiently. The direct statement definitions brought forth, as a result of consensus between the group members were:

| Problem Solving | Analyses available information and identifies key issues objectively; |
|                 | Identifies and understands problems globally rather than isolated bits; |
|                 | Generates decision criteria in order to fins solutions for the problems. |

The second competency considered was “ Team-work” which refers to the capacity for working well in a team and for actively contributing to the solution of problems. The direct statement definitions are the following:

| Team-work | Acknowledges individual differences in others, values these differences and adjusts them in order to reach identified goals; |
|           | Puts team goals before own; |
|           | Seeks to establish win-win solutions. |

The third competency considered was “Self and Subordinate development” which refers to the capacity for going beyond routine demands. The following direct statement definitions were brought forth:

| Self and Subordinate Development | Encourages subordinates development; |
|                                  | Continuously assigns subordinates activities and responsibilities that build their capabilities; |
|                                  | Encourages training and knowledge sharing; |
|                                  | Demands subordinate commitment to organizational values and codes of conduct. |

The fourth competency considered was “Discipline” which refers to the military condition of permanent commitment to duty and self-sacrifice. The following direct statement definitions were identified:

| Discipline | Acts according to organizational norms, rules and regulations; |
|           | Displays authority in an efficient and sober manner; |
|           | Instils others to act according to organizational norms, rules and regulations. |

At the end of this work session the WG manifested its satisfaction with the identified model.
The next steps are a simulation, concerning the identification of a response scale and the evaluation procedure that should follow.

At this point, a response scale for the appraisal of each direct statement definition would have to be identified. There are various alternatives namely frequency type scales, agreement type scales, numeric scales, etc, but there are no real certainties regarding the best type of response scale. To simplify we will use a yes/no dichotomic scale. In this manner, the evaluator would be presented with each direct statement definition, and asked to say whether or not (yes or no) it represented the performance observed during the observation/evaluation period.

Each competency cluster will have to be analysed and evaluated in order to determine its weight in the overall appraisal model considering if the direct statement definitions (DSD) is “very important” or “important”, where “very important” would be perceived as being twice as important as the “important” statements. Thus, a negative grade (“no”) on “very important” statements would be twice as damaging, whilst a “yes” grade would have the same value as the remaining.

We may exemplify by taking into consideration the second scenario where “very important” statements are considered twice as important as “important” statements, with relation to a negative grade (score). In this case the following reference levels would be taken into concern:

<table>
<thead>
<tr>
<th>Maximum level</th>
<th>ALL “Yes” answers in “very important” as well as in “important” direct statement definitions (DSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superior Level</td>
<td>“Yes” answers in ALL “very important” DSD and ONE “Yes” answer in an “important” DSD’s per competency</td>
</tr>
<tr>
<td>Neutral Level (+)</td>
<td>“Yes” answers in ALL “very important” DSD and only ONE “No” answer in an “important” DSD per competency</td>
</tr>
<tr>
<td>Neutral Level (-)</td>
<td>“Yes” answers in ALL “very important” DSD and “No” answers in ALL “important” DSD per competency</td>
</tr>
<tr>
<td>Mediocre Level</td>
<td>“No” in ALL “very important” DSD and “Yes” answers in ALL “important” DSD per competency</td>
</tr>
</tbody>
</table>
These proceedings would enable the identification of performance intervals into which all evaluated officer could be grouped, for example, weak, below average, average, superior and excellent performer.

3. Model Building

The Linear additive value model

At this point we will have to define preference relations between the evaluation factors – competencies, in order to produce an ordinal listing, only in this way, will we be able to aggregate the evaluation factors and determine an overall performance indicator.

The linear additive value model is a simple Multiple Criteria Decision Analysis model. Its construction should be done using the MACBETH approach (Measuring Attractiveness by a Categorical Based Evaluation Technique – cf Bana e Costa and Vansnick, 1997 and 1999) – an interactive approach that guides the construction of a value scale, based on the qualitative value judgments.

In this case the following activities will have to be put in practice.
- Assign weighting constants to the plausible ranges of the descriptors, in order to harmonise and aggregate the attractiveness related to each competency;
- Perform feedback and sensitivity analyses in order to build a requisite model.
In our case the additive formulation that would be adopted would be:

\[ V(o) = \sum_{j=1}^{14} k_j \cdot v_j(o) \]

Where,

- \( V(o) \) – is the global value of a fictitious officer \( o \),
- \( v_j(o) \) – is the partial value of a fictitious officer \( o \) on the \( j \)-th competency,
- \( k_j \) – is the weight of the \( j \)-th competency,

In this case,

For \( 0 \leq v_j(o) \leq 100 \), \( v_j \) and \( 0 \leq V(o) \leq 100 \)

\[ \sum_{j=1}^{14} k_j > 0 \quad (j = 1, \ldots, 14) \]

**Determine the weights with MACBETH**

In order to conduct this procedure the facilitators will have to ask the WG to consider all fourteen competencies at their worst level, and then choose only one to pass from the worst to the best level. Doing this successively with all the other competencies we obtain an ordered list of competencies.

As result of the weighting session an ordered list of evaluation factors would be reached:
After this, the WG will have to qualitatively evaluate the difference of attractiveness between each pair of competencies regarding the six MACBETH semantic categories “very weak, weak, moderate, strong, very strong, extreme”. If there is some inconsistency judgement MACBETH detects this situation and presents suggestions to correct the judgements.

This result of this procedure would be a of a similar nature as the image bellow:
This matrix represents the difference of attractiveness between the fourteen evaluation factors. As a result the overall weighting would be as follows:

![Weights Chart](chart.png)

We can then proceed to join up each direct statement definition with its respective cluster in order to evaluate the overall consistency of the model. In other words we would be able to present the overall weight actually made by the evaluators, for example, as a result of the weights we determined we would find the following scenario: The competency cluster with the most relative importance in the performance appraisal of the Army officers would be – Institutional, followed by Technical, Interpersonal and finally Personal competencies.

<table>
<thead>
<tr>
<th>Institutional</th>
<th>34.1</th>
<th>Technical</th>
<th>23.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loyalty</td>
<td>+ (10.9)</td>
<td>Decision Making Process</td>
<td>+ (8.5)</td>
</tr>
<tr>
<td>Discipline</td>
<td>+ (9.0)</td>
<td>Planning and organization</td>
<td>+ (8.3)</td>
</tr>
<tr>
<td>Readiness to serve</td>
<td>+ (7.4)</td>
<td>Problem Solving</td>
<td>+ (6.3)</td>
</tr>
<tr>
<td>Courage</td>
<td>+ (6.8)</td>
<td>Professional Expertise</td>
<td>+ (0.1)</td>
</tr>
<tr>
<td><strong>Interpersonal</strong></td>
<td><strong>26.2</strong></td>
<td><strong>Personal</strong></td>
<td><strong>16.5</strong></td>
</tr>
<tr>
<td>Leadership</td>
<td>+ (10.8)</td>
<td>Determination</td>
<td>+ (7.3)</td>
</tr>
<tr>
<td>Team work</td>
<td>+ (7.8)</td>
<td>Self and subordinate development</td>
<td>+ (6.1)</td>
</tr>
</tbody>
</table>
As we can see, the data is consonant with the initial generalized feeling regarding the relative importance of each competency cluster in the overall evaluation scheme.

The application of the reference rules already presented with the identified weights will give an output of reference levels as we can see in the figure below.

The overall reference scale is consonant with Kahneman & Tversky’s conception of decision-making when in presence of risk. These Nobel Prize of Economy winners consider that people have a natural disposition to avoid risk; this is relevant in performance appraisal if we take consequential actions of a negative evaluation into consideration, so, when a negative answer is given it should be seriously taken into account. In the model we exemplify with, an option for a “no” answer regarding the performance of an officer in a particular competency is, therefore, considered twice as “penalizing”.

These reference levels are not regarding actual scores, they would simply indicate if an officer would need more or less corrective interventions, such as training and development, coaching, etc. As in Tversky and Kahneman’s function the interval between scale points on the negative side are bigger than the ones encountered on positive evaluations.

Finally, if we consider these weights to be accurate we would then be able to pass on to the final step – the elaboration of recommendations that is the final phase of analysis. In order to evaluate the global consistency of the identified competency model, we recommend a sensitivity analysis.

4. Final Comment

The intervention process described in this paper enabled the Work Group to accomplish a “shared understanding of the issues and a sense of common purpose as well as a mutual commitment to action” (Phillips and Phillips, 1993). Its objective was achieved with the identification of key performance concerns to be taken as evaluation criteria in the appraisal of army officers.

The process consultation interventions used by the facilitators have proved to be valuable throughout the work sessions, as they have consistently contributed to help the participants to take the problem, its analysis and resolution into their own hands, without
passing on this responsibility on to the facilitators, fostering ownership for the resolutions and their implementation.

The use of the MACBETH approach is an adequate and easy to use alternative to evaluate the global consistency of the identified competency model. In this way, decision makers may reassure the validity of the practical results (outputs) the model will encompass.

Finally, regarding the complexity and multidimensionality of this particular problem, we propose a multimethodology framework (Mingers & Gill, 1997), in which the combination of parts of methodologies, tools, techniques, etc, is an efficient alternative when tackling complex problems where facilitators need to understand every stage of the decision process and choose which approach will produce the best results, especially when we consider methodologies as repertoires and not recipes. The dynamic world calls for flexible and dynamic approaches. Problems are indeed messy.

In our approach, we enriched the multiple criteria decision analysis methodology with contributions borrowed from other decision support & group support methodologies, under a single paradigm – the learning paradigm – that guaranteed the theoretical and practical cohesion and consistency of the process at all stages of development.

REFERENCES


SITUATIONAL JUDGMENT TEST RESEARCH CONTEXT

Tonia S. Heffner, Ph.D.
Senior Research Psychologist
U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Ave, TAPC-ARI-RS, Alexandria, VA 22333
heffnert@ari.army.mil

Deirdre J. Knapp, Ph.D.
Manager Assessment Research and Analysis Program
Human Resources Research Organization
66 Canal Center Plaza, Alexandria, VA 22314
dknapp@humrro.org

INTRODUCTION

The U.S. Army Research Institute for Behavioral and Social Sciences (ARI) conducted research on the junior Noncommissioned Officer (NCO) promotion system – promotion from specialist/corporal (E4) to sergeant (E5) and sergeant (E5) to staff sergeant (E6). This research, Maximizing 21st Century NCO Performance (NCO21), was initiated in FY99 to examine the role of the NCO in the next quarter century, 2000-2025. The three primary research objectives were to (a) identify the performance requirements and the associated knowledges, skills, and attributes (KSAs) NCOs would need to succeed in the changing military environment; (b) identify criterion and predictor measures to assess the performance requirements and KSAS; and (c) validate the predictors in relation to the criterion measures and to examine the incremental prediction of performance provided by the future-oriented predictor measures (Ford, R. Campbell, J. Campbell, Knapp, & Walker, 2000; Knapp et al, 2002; Knapp, McCloy, & Heffner, 2002).

The current junior NCO promotion system considers four administrative factors when evaluating a soldier: (a) civilian education; (b) military education; (c) awards, decorations, and achievements; and (d) military training, as well as the commander’s recommendation and the promotion board appearance. An early task of this research was to determine if all of the critical components of current NCO performance, as well as those predicted to be important for Objective Force soldiers, were being tapped by the current system. To accomplish this task, a set of critical KSAs were identified with the assistance of military and personnel selection subject matter experts. This task revealed that several KSAs anticipated to be important for NCO performance were overlooked in the current promotion factors. The situational judgment test (SJT) was one of a variety of potential performance predictors developed for the NCO21 project to assess these critical KSAs. A substantial portion of the validation effort was to determine how well the predictors contributed to the prediction of performance above and beyond the administrative predictors included in the current promotion system.
SJT examinees are presented with job relevant situations and asked to select, from a list of possible alternatives, the best action to take in this scenario (see Figure 1). SJTs have a history as a potential predictor measure with the U.S. Army. During the 1980s, an SJT was used as a criterion measure in the Project A/Career Force research program to assess a variety of constructs, including supervisory and leadership requirements (J. Campbell & Knapp, 2001). An SJT was also successfully used as a predictor in an earlier Army project to examine the NCO promotion system—Expanding the Concept of Quality in Personnel (ECQUIP; Peterson et al., 1997). Outside of the selection arena, SJTs have been used for leadership development (Platoon Leader Questionnaire [PLQ]; Hedlund et al., 1999). Tests with components similar to SJTs have been used in other Army contexts (e.g., Legree, Heffner, Psotka, Martin, & Medsker, 2002).

The NCO21 SJT was one of several predictor measures hypothesized to increase the prediction of junior NCO performance beyond the administrative predictors included in the current promotion system.

**METHOD**

The research reported in this paper and the associated papers (Putka, Waugh, & Knapp, 2002; Waugh, Putka, & Sager, 2002) concentrates on the development and validation efforts for the situational judgment test (SJT) and the other predictors. During the development and validation, a unique relationship between the SJT measure and pay grade was identified and explored in detail. Only the relevant predictor and criterion measures will be described. A full description of the research initiative can be found in Ford, R. Campbell, J. Campbell, Knapp, and Walker (2000), Knapp, Burnfield, et al. (2002), and Knapp, McCloy, and Heffner (2002).

**Predictor Measures**

The *Experiences and Activities Record (ExAct)* was intended to identify experiences and activities that contribute to improved junior NCO performance. The assessment tool had 46 self-report items that measured how frequently junior soldiers engaged in tasks such as training others, acting as supervisors, and working with computers. Based on a factor analysis of the item responses, three scores were derived from the ExAct: Supervisory Experience, General Experience, and Computer Experience (see Table 1). Internal estimates of reliability were acceptable.

Figure 1. Sample situational judgment test item.

<table>
<thead>
<tr>
<th>Reliability Estimates</th>
<th>E5</th>
<th>E6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Situational Judgment Test</strong>&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Directing, Monitoring and Supervising Individual Subordinates</td>
<td>.73</td>
<td>.68</td>
</tr>
<tr>
<td>Training Others</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team Leadership</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concern for Soldiers’ Quality of Life</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cultural Tolerance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivating, Leading, and Supporting Individual Subordinates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relating to and Supporting Peers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem Solving/Decision Making Skill</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Experience and Activities Record</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General Experience</td>
<td>.82</td>
<td>.80</td>
</tr>
<tr>
<td>Supervisory Experience</td>
<td>.84</td>
<td>.82</td>
</tr>
<tr>
<td>Computer Experience</td>
<td>.82</td>
<td>.77</td>
</tr>
<tr>
<td><strong>Assessment of Individual Motivation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dependability</td>
<td>.57</td>
<td>.55</td>
</tr>
<tr>
<td>Adjustment</td>
<td>.69</td>
<td>.70</td>
</tr>
<tr>
<td>Work Orientation</td>
<td>.73</td>
<td>.69</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>.70</td>
<td>.70</td>
</tr>
<tr>
<td>Physical Conditioning</td>
<td>.64</td>
<td>.64</td>
</tr>
<tr>
<td>Leadership</td>
<td>.65</td>
<td>.61</td>
</tr>
<tr>
<td><strong>Biographical Information Questionnaire</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tolerance for Ambiguity</td>
<td>.52</td>
<td>.34</td>
</tr>
<tr>
<td>Openness</td>
<td>.68</td>
<td>.62</td>
</tr>
<tr>
<td>Hostility to Authority</td>
<td>.72</td>
<td>.71</td>
</tr>
<tr>
<td>Manipulativeness</td>
<td>.77</td>
<td>.75</td>
</tr>
<tr>
<td>Social Maturity</td>
<td>.69</td>
<td>.67</td>
</tr>
<tr>
<td>Social Perceptiveness</td>
<td>.83</td>
<td>.83</td>
</tr>
<tr>
<td>Interpersonal Skill</td>
<td>.52</td>
<td>.56</td>
</tr>
<tr>
<td>Emergent Leadership</td>
<td>.82</td>
<td>.80</td>
</tr>
</tbody>
</table>

Note. <sup>a</sup> The SJT internal estimate was computed for the 24-item test.

The Situational Judgment Test (SJT) was intended to measure eight KSAs including leadership and decision making (see Table 1) by presenting several brief scenarios per KSA and asking the soldier to identify the best and worst possible actions from a list of four alternatives. Some of the 40 test items were adapted from other Army measures (e.g., Project A [J. Campbell & Knapp, 2001], ECQUIP [Peterson et al., 1997], PLQ [Hedlund et al., 1999]) and others were developed specifically for this research. The items were scored in relation to item responses provided by SMEs. Following the
concurrent validation, shorter 24-item forms of the SJT were constructed to reflect the
need to reduce test length as much as possible to meet operational requirements. The 24-
item forms are somewhat different for the E5 and E6 pay grades, whereas the original 40-
item form was the same for both pay grades.

A second situational judgment test (SJT-X) was intended to measure the futuristic
KSA “knowledge of interrelatedness of units” using three lengthier scenarios. These
scenarios were developed explicitly for this research based on themes identified as
important for the future (Ford et al., 2000). Like the SJT, the items were scored in
relation to item responses provided by SMEs. (Additional SJT information is available in
Putka et al. [2002] and Waugh et al. [2002].)

The Assessment of Individual Motivation (AIM; White & Young, 1998) was
intended to assess six temperament qualities (see Table 1). The AIM is a 38-item quas-
ipsative assessment tool that requires soldiers to select, from a list of four alternatives, the
temperament statement that most and least describes them. This forced-choice technique
is used to reduce the impact of faking. The U.S. Army, as part of an operational test
program, is administering a version of the AIM. Internal consistency estimates of
reliability were marginally acceptable for most scales. (Additional AIM information is
available in White [2002].)

Similarly, the Biographical Information Questionnaire (BIQ; Kilcullen, Mael,
Goodwin, & Zazanis, 1996) assessed eight characteristics (see Table 1) by asking soldiers
to answer questions about their attitudes and past experiences. The BIQ is a 154-item
multiple choice format measure developed from several U.S. Army operational measures.
Internal consistency estimates of reliability were acceptable for most scales. (Additional
BIQ information is available in Kilcullen [2002].)

Criterion Measure

The NCO21 project included two ratings-based criterion measures—a ratings
instrument pertaining to current performance and another instrument that asked
supervisors to predict how well their soldiers would perform under six future-oriented
scenarios. The criterion measure used in the SJT analyses reported in this symposium was
the current performance instrument. Using the information gleaned from the early stages
of the research, 27 performance requirements were identified and a behaviorally anchored
rating scale was constructed for each. The performance assessment instrument consisted
of a statement of the research dimension; a definition of the dimension; a 7-point rating
scale with global indicators of high, moderate, and low performance; and at least three
examples of typical performance at for the three global rating levels. In addition to the 27
performance requirements, supervisors also rated their soldiers on Overall Performance
and Senior NCO Potential. Based on the field test results, the 27 performance
requirements were collapsed into 19. The primary criterion score was the average of the
19 individual scale ratings.
Data Collection

The measures were field tested on 513 E4, E5, and E6 soldiers at three Army installations. Job performance ratings data were collected from supervisors of the E5 and E6 field test participants. The field test was critical for evaluating and refining the instruments and also for trying out the data collection protocols and database management procedures.

We used a concurrent validation design in which we collected predictor data on E4, E5, and E6 soldiers and criterion data from E5 and E6 soldiers. The E4 predictor data were used as a basis for correcting validity estimates for range restriction. Data were collected at seven sites from approximately 1,900 soldiers and 988 supervisors representing 122 MOS. Table 2 shows the sample sizes by pay grade and location. The goal was to collect ratings from two supervisors per E5/E6 participant. Roughly 70% of the soldiers had at least one supervisor rater; only about 30% had two raters.

Table 2.
Validation Sample Sizes

<table>
<thead>
<tr>
<th>Grade</th>
<th>Hood</th>
<th>Bragg</th>
<th>Lewis</th>
<th>Riley</th>
<th>Campbell</th>
<th>Carson</th>
<th>Stewart</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>E4</td>
<td>82</td>
<td>62</td>
<td>56</td>
<td>36</td>
<td>89</td>
<td>59</td>
<td>65</td>
<td>449</td>
</tr>
<tr>
<td>E5</td>
<td>177</td>
<td>162</td>
<td>69</td>
<td>143</td>
<td>126</td>
<td>102</td>
<td>106</td>
<td>885</td>
</tr>
<tr>
<td>E6</td>
<td>135</td>
<td>58</td>
<td>67</td>
<td>68</td>
<td>91</td>
<td>75</td>
<td>63</td>
<td>557</td>
</tr>
<tr>
<td>Total</td>
<td>394</td>
<td>282</td>
<td>192</td>
<td>247</td>
<td>306</td>
<td>236</td>
<td>234</td>
<td>1,891</td>
</tr>
</tbody>
</table>

Analyses

The concurrent validation database was analyzed to (a) examine the psychometric characteristics of the predictor and criterion measures, (b) estimate the criterion-related validity of each predictor score, and (c) estimate the incremental validity of the experimental predictors over the current administrative information used as a basis for NCO promotions. Overall, the results were very promising. Along with several other predictor scores, the SJT composite score showed incremental validity over the current system, but the validity was much stronger for E5 soldiers compared with E6 soldiers.

CONCLUSIONS

The NCO21 research effort, in more detailed results to be discussed in the remainder of this symposium and in Knapp, McCloy, et al. (2002), shows the SJT to be (a) a promising predictor that should be relatively easy to implement in a large-scale assessment program that (b) has several complex characteristics that we are just
beginning to understand. As it has in other settings (see McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001 meta-analysis), the NCO21 SJT was an effective predictor of performance – in this case both current and expected future performance. In addition to their criterion-related validity, well-constructed situational judgment tests are desirable for their inherent content and face validity. It is also easier to develop a discriminating situational judgment test than one might think. Moreover, operational SJTs can be managed in much the same way as other large-scale traditional test programs (e.g., with large item banks that can be expanded through transparent administration of experimental items on operational tests).

But situational judgment tests, perhaps particularly the one developed in NCO21, show increasing complexity the more they are studied. For example, researchers have tried a variety of scoring schemes, but it is not clear that the same algorithm is the best across applications and it seems likely that different algorithms represent somewhat different underlying constructs (e.g., picking the right action versus avoiding the worst actions). A particular mystery in the NCO21 research is the criterion-related validity differences across pay grades. The obvious answers (e.g., range restriction) do not explain this finding, but the follow-up analyses (Putka et al., 2002; Waugh et al., 2002) we have conducted to try to understand the differences has yielded a much better understanding of the response styles underlying the options provided on the test, as well as more questions about procedures for developing scoring keys (e.g., is the “right” answer dependent on the respondent’s organizational role). This, in turn, points out important factors that should be considered when developing alternate forms. Development of alternate situational judgment test forms is an area that has received relatively little research attention and which promises more challenges for future research. These are challenges of the good kind, however, and we are confident that the NCO21 SJT and similar measures for other applications have an important place in U.S. Army personnel research.

REFERENCES


Does Experience change Perceived Team Effectiveness in Peace Support Operations?

Prof. Dr MSc Jacques Mylle
Department of Psychology
Royal Military Academy
Brussels – Belgium

Aside the improvement of the capabilities of the individual, team effectiveness has always been a center of interest. In former times the attention was exclusively focused on the internal aspects of the team such as cohesion. In the nineties attention shifted towards an external perspective, taking into account situational aspects.

A lot of tasks during peace support operations (PSO) have to be performed by small military units (say from 4 to 10 people). These PSO tasks are characterized by the necessity of teamwork and a high level of autonomy due to the large distances between the (sub)unit(s) executing a task and their superior, and the potentially quickly evolving situation as well, called turbulent situations. A particular parameter in such a context is “boundary management”; i.e. the activities developed at the “border” between the internal and the external world of the team considered. These boundary activities are deployed to handle the turbulence in the situation. PSO can be considered as a turbulent situation and, hence, it is interesting to verify how the team deployed evaluates its own effectiveness based on a number of given parameters related to autonomy and boundary management.

Method

We are conducting a study together with co-workers of the Department of Psychology at Catholic University of Leuven, aiming at measuring the “perceived team effectiveness in turbulent situations”. This study is intended both as cross sectional research allowing comparisons between several types of units deployed and as longitudinal research, measuring the changes in perceived efficacy before and after completion of a mission. Based on a model set up for this purpose, a questionnaire has been elaborated aiming at measuring four variables; i.e. Turbulence, Autonomy, Boundary Management and Effectiveness. Each of those variables is broken down in several facets, which in turn are measured by a scale containing a variable number of items. Turbulence is measured at three “levels”: task – and job related turbulence, organizational turbulence and extra-organizational turbulence. Autonomy is evaluated on 11 facets; among distribution of responsibilities, needed equipment and regulation of workload. The subscales of boundary management correspond to the four roles as defined by Ancona; i.e. searching for information (scout), profiling the team (ambassador), co-ordinate with other teams having a similar task (co-ordinator) and protecting the team from external interference (guard). Because these roles require a lot of communication we added a scale for this

---

40 For former presentations on this topic, see IAMPS 2001 and 2002 and IMTA 2001.
issue. Finally to better understand why the deployed some boundary activities we asked for their motives to do so.

Effectiveness is measured with respect to five types of outcome: product, procedures, process, viability and co-ordination with other teams.

All items are evaluated on a 5-point Likert scale, going from “not applicable” to “entirely” applicable” or from “never” to “always”.

**Aim**

The central question in this paper is to verify to what extent the experience of a deployment changes the perception of turbulence, autonomy, boundary management and effectiveness as compared to the perception at the end of the training before deployment.

**Method**

Data concerning “before deployment” have been collected after an intense training period, 14 days before deployment in a sample of about 200 soldiers out of a 800 belonging to the infantry, armor troops, pioneers, logistic support and medical support. They are part of a battalion size task force Belukos VI, which was deployed from the beginning of April 2001 until the beginning of August 2001. Data concerning “after deployment” have been collected at the end of the year, after the units “recovered” from the deployment. Due to circumstances the second wave sample consists of 172 persons only belonging to Infantry, Pioneers and Medical Support. Ninety of them have completed the questionnaire in both waves.

Given that most of the respondents did not remember their identification code of the first wave, it was not possible to use a paired sample t-test to answer our question stated above. In other words we had to test our hypothesis at an aggregated level and as if the sample of wave 1 was independent of the sample in wave 2.

**Results**

**Differences in perceived turbulence**

The Turbulence scale consists of three subscales measuring three “levels”; i.e. task turbulence, organizational turbulence and extra-organizational turbulence. The means before deployment of these subscales are respectively 3.37, 2.71, 3.29 and 3.12 for the overall turbulence. This means slightly above the average score of 3 whereas the means after deployment are all a little bit lower with 3.25, 2.67, 3.22 and 3.05 respectively.

At the task level we observe a significant different with respect to “task complexity” (? = -0.33, p = .000). This means that the respondents experienced the real tasks once deployed as less complex than the same tasks in the training situation. Although the difference of means on the other subscales are not significant, the difference on the subscale “task” is significant (?= -0.12, p= .09).

At the level of the organization, the respondents perceived the organizational environment as more stable during the deployment as during the training (? = -0.19, p =
.03). All other aspects do not show any significant difference and the subscale as a whole
neither.

The change in perceived extra-organizational turbulence is not significant.

Thus, taken all together, the differences in environmental turbulence are not significant.

In other words, there are no noticeable differences between the expectations based on
training and the experienced turbulence in the operation zone.

**Differences in perceived autonomy**

The overall estimation of autonomy of nearby 2.3 before deployment decreases to 2.15
after deployment, if we do not control for turbulence. In this case, there are only two
facets that lead to significant differences; i.e. “responsibilities” (\( \beta = -0.45, p=.03 \)) and
“rewarding” (\( \beta = -0.42, p=.00 \)).

On the contrary, if we consider only the upper 33% of the sample of the turbulence
distribution (condition turbulence “high”) we find much more differences, all pointing in
the negative direction: procedures (\( \beta = -.23, p =.07 \)), Responsibilities (\( \beta = -.28, p=.03 \)),
Control (\( \beta = -.17, p =.09 \)), training (\( \beta = -.23, p =.03 \)), rewards (\( \beta = -.35, p =.00 \)) and
workload (\( \beta = -.15, p =.03 \)). Taken all together, the overall difference in autonomy is
also significant (\( \beta=-.20, p=.01 \)).

Thus, the expectations with respect to autonomy were consistently higher than the actual
degree of autonomy they had during deployment but the differences are mostly not
significant.

Precisely having less autonomy when required by the high degree turbulence may have a
negative impact on motivation and thus perhaps on effectiveness.

**Boundary Management**

Former research has shown that boundary management acts as buffer for turbulence.
Stated otherwise, if there is no environmental turbulence, there is no need for boundary
management; moreover boundary management activities may be counterproductive in
this case.

Not controlling for turbulence results in non-significant differences for all aspects of
boundary management. If we use only the 33% of the sample that reports high perceived
turbulence, we observe any significant difference, except for the guard subscale guard, as
far as this scale is based on the highest loadings on that factor (\( \beta = +.39, p =.01 \)). In
other words, the respondents tend more to protect the team from external interference as
they did on training before deployment. Nevertheless, the observed difference with
respect to the guard role seems to be an effect of operationalization when comparing the
non significant results when the roles are computed on the base of all loadings higher
than .40 versus computing scales on the basis of the highest loading per factor

It is not surprising that we found no significant differences in boundary management
because, on average, the respondents did not perceive any difference with respect to
turbulence.
Effectiveness
The values of perceived overall effectiveness are average (3.08 – 3.15). Given that we made a distinction between “all degrees of turbulence” and “high turbulence” for the other variables, we need to do the same with respect to the effectiveness scales. Not controlling for turbulence, results in no significant differences for the subscales “product” nor “procedure” but we observe a significant difference at the level of the process; i.e. the interpersonal relationships (? = .30, p = .00). This difference can be explained by the difference on the subscales “participation” (? = +.62, p = .00) and “openness” (? = +0.52, p = .00) respectively.

When controlled for turbulence the overall effectiveness is a little bit but not significantly lower (3.15 vs 3.09 before deployment and 3.15 vs 3.09 after deployment). The same scales still show a significant difference; i.e. for participation (? = .71, p = .00) and openness (? = .49, p = .00).

Thus the soldiers believe they have the same level of product effectiveness before and after deployment. The process effectiveness increased substantially at the interpersonal level but not at the team level because the results for cohesion did not change (? 3.6).

Discussion and conclusions
Turbulence is estimated to be average and “manageable during the training period and during the deployment as well. The training situation and the exercises were thus sufficiently realistic. Conversely, if the task force commander sets up realistic training, and every else is kept constant, the soldiers will be effective in the operations zone. To that extent we can speak of an anticipated effectiveness, which a motivating factor, because related to an intern locus of control (see Bandura’s Self-efficacy Theory also).

The degree of autonomy seems to pose a problem. A mean score around 2.3 means that the perceived degree of autonomy is below average. This is by itself perhaps “reasonable” in a hierarchical structure as a task force. But the fact that they experienced less autonomy in the operation zone as compared to their expectations created in training may have a negative impact on their perceived and actual effectiveness. Moreover, this misfit increases with increasing turbulence. This is precisely the opposite of what is needed: the more turbulence, the more autonomy.

Boundary management is also estimated below average (? 2.6) but this may be due to the fact that this concept is new for military team leaders. Dealing explicitly with the boundary management roles in training will probably increase the competencies at that level, contribute to deal better with the environmental turbulence and thus be more effective.

Finally, the product effectiveness is estimated average too (?3.1). At the level of the military proficiency, in terms of “product” and procedures, the experiences equal the expectations. This means that the training was effective with respect to these issues. Aside, there is a significant intensification of the interpersonal relationship during the deployment. This change can be explained by the fact that a task force is composed of different units coming from different garrisons, and even within a unit, people from
different platoons and squads are regrouped in such a way to compose full manned groups. Training before deployment is partly conceived as team building. However the data show that this process is still in progress during the deployment.

As an overall conclusion, we can say that our soldiers feel effective in both situations.

It would be worthwhile to verify if there are differences between soldiers with a first deployment and those with more than one deployment. But breaking down further the sample will pose a problem of power of the test due to the limited number of respondents in that group.

Finally, it must be stressed that the results obtained on the basis of the current questionnaire have to be interpreted with caution because to be seen as a pilot study and a first phase in the validation process of a new instrument.
SIMULATION TESTBED FOR THE EVALUATION OF PERFORMANCE (STEP): INNOVATIVE CRITERIA

Kathleen P. Hess, Ph.D.
Industrial/Organizational Psychologist
Aptima, Inc., 12 Gill Street, Suite 1400, Woburn, MA 01801
khess@aptima.com

Elliot E. Entin, Ph.D.
Senior Scientist, Experimental Psychologist
Aptima, Inc., 12 Gill Street, Suite 1400, Woburn, MA 01801
entin@aptima.com

Peter Miller, B.S.
Operations Analyst and Simulation Engineer
Aptima, Inc., 12 Gill Street, Suite 1400, Woburn, MA 01801
pmiller@aptima.com

As part of the team collecting validity evidence for the future-oriented junior NCO promotion system, Aptima’s role was to develop a nontraditional way to measure job performance. We developed the STEP tool – a simulation-based performance assessment environment constructed using the Distributed Dynamic Decision-making (DDD) simulation development tool.

The DDD, a human-in-the-loop simulation that flexes important cognitive processes and tasks, has been a major component of an individual and team performance research program that has been underway for almost 15 years. Co-developed by Aptima’s President and CEO, Daniel Serfaty (Kleinman & Serfaty, 1989; Kleinman, Young, & Higgins, 1996; Serfaty & Kleinman, 1985), the DDD is a unique software toolset and computer system developed to study issues such as situation assessment, decision making, and information management.

In developing STEP, Aptima used its extensive experience in creating synthetic tasks for performance measurement in military environments. Aptima’s abstract and configurable DDD simulation provided a powerful, flexible development environment for creating synthetic tasks. The remainder of this paper describes Aptima’s contribution to the validation process.

METHOD

Participants
Seventy-four Sergeants and Staff Sergeants from three Army installations participated in the research.

41Findings developed with Army Research Institute Support, Contract #DASW01-00-C-3012. The view, opinions, and/or findings contained in this report are those of the authors and should not be construed as an official Department of the Army position, policy, or decision.
Apparatus and Materials

STEP Tool

STEP is a tool capable of simulating future, technologically-intense jobs so that performance data can be collected when humans perform in the simulated environment. The scenario developed for this project takes place in an urban setting. The town has buildings, roads, and impassable hills (see Figure 1 for a screenshot). Terrorist activities have resulted in chemically contaminated areas and the presence of unexploded ordnances (UXOs). All of the residents fled the area at the onset of the terrorist activities and are about to return. Some of them are wounded, some need food, some are friendly to the allied forces, and others are hostile. The allied forces must ensure the safety of returning residents and future forces. The allied forces also help the refugees when they are wounded or hungry and detain them when they are hostile. This computer simulation was developed for two participants – a leader and a subordinate (note that during data collection the subordinate was actually a confederate who made (preplanned) mistakes and needed guidance). Each participant controls two High Mobility Multipurpose Wheeled Vehicles (HMMWVs) with 4 (simulated) soldiers in each. Each simulated soldier has special skills needed to perform specific tasks. For example, one soldier under the leader-participant’s control possesses the majority of expertise necessary to clear chemicals whereas the subordinate-participant controls the soldier possessing the majority of expertise needed to defuse UXOs.

Measures and Instruments

Aptima Criterion Measures

Using the STEP tool, we collected data needed to provide criterion measures for the NCO21 research program. Measures were derived from three sources: the participant, observers, and the STEP tool. Participants completed post-scenario questionnaires designed to assess their perception of personal and team performance and their thoughts on how their actions influenced other units. Observers rated the participants when they briefed their subordinate on the mission and when they gave midterm and post-scenario oral situation reports (briefs) using behaviorally anchored rating scales. STEP’s log and history files provided performance data measuring how well the participant followed procedures and the total number of tasks handled successfully.
**Use of the Measures.** STEP was designed to simulate performance requirements anticipated to be important for future NCOs. The measures were used to quantify the participant’s performance on 9 performance components. Each component was measured by more than one source (see Table 1). As a result, before the validation could be completed, the appropriate item scores were combined to create the appropriate composite scores. These composite scores were then combined with other criterion scores collected through more traditional means to complete the selection test validation.

Table 1. Performance Measures by Source

<table>
<thead>
<tr>
<th>Performance Component</th>
<th>Abbreviation</th>
<th>Source</th>
<th>Computer</th>
<th>Participant Survey</th>
<th>Observer</th>
<th>Confederate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Skills</td>
<td>comp</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Oral Communication</td>
<td>oral</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Adaptability</td>
<td>adapt</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Safety Consciousness and Adherence to regulations,</td>
<td>safe</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>policies, and procedures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leadership</td>
<td>lead</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Knowledge of the interrelatedness of units</td>
<td>know</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Management and coordination of multiple battlefield functions</td>
<td>mgmt</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Problem Solving and Decision Making</td>
<td>prob</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Information Management</td>
<td>info</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

**HumRRO Measures**

*Criteria.* As described in Heffner and Knapp (2002), two performance rating criterion measures were developed: (a) Observed Performance Rating (OPR) Scales and (b) Expected Future Performance Rating (EFP) Scales.

*Predictors.* Scores from five instruments were examined by HumRRO as predictors in the current investigation: the NCO21 Situational Judgment Test (SJT), Simulated Promotion Point Worksheet (SimPPW), the Assessment of Individual Motivation (AIM), the Biographical Information Questionnaire (BIQ), and the Armed Services Vocational Aptitude Battery General Technical Score (ASVAB GT). These instruments are described in more detail in Heffner and Knapp (2002).

**Procedure**

All participants received the same training and background information in preparation for use of the STEP tool. Their mission entailed: 1) assessing civilians’ medical and food related needs and providing only the food/medicine they need; 2) providing civilians with a safe passage west; 3) identifying and managing/cleaning chemical areas; 4) identifying and managing/neutralizing UXOs; 5) identifying and detaining hostiles; and 6) protecting civilians from UXOs and chemical sites. In addition to completing the STEP performance criterion simulation, all soldiers were administered
RESULTS

Interrater Reliability

For three of the assessments – Post Brief Questionnaire, Post Game Performance Assessment Rating Form, and Hotwash Performance Assessment Rating Form – both the observer and the subordinate (confederate) rendered parallel but independent ratings. Thus, we are able to examine the inter-rater reliability for these three instruments. The correlations between the observer and the subordinate ratings for the three instruments, by data collection site are shown in Table 2. The Post Brief Questionnaire and Hotwash Performance Assessment Rating Form correlations show moderate inter-rater relationship for each of the data collection sites. The primary observer base performance assessment, Post Game Performance Assessment Rating Form, however, exhibits moderate-high to high inter-rater correlations for each of the samples. This leads us to conclude that the assessment of important performance requirements were done in a consistent and reliable manner.

Table 2. Inter-rater correlations for three assessment instruments by data collection site

<table>
<thead>
<tr>
<th>Instrument</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post Brief Questionnaire</td>
<td>0.50</td>
<td>0.52</td>
<td>0.46</td>
<td>0.55</td>
<td>0.55</td>
<td>0.49</td>
</tr>
<tr>
<td>Post Game Performance Assessment Rating Form</td>
<td>0.76</td>
<td>0.73</td>
<td>0.64</td>
<td>0.80</td>
<td>0.77</td>
<td>0.67</td>
</tr>
<tr>
<td>Hotwash Performance Assessment Rating Form</td>
<td>0.48</td>
<td>0.73</td>
<td>0.60</td>
<td>0.52</td>
<td>0.77</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Aptima’s Constructs

Each item comprising the nine instruments completed by the observers and the three instruments derived from the participants (with some SME scoring) was designed to assess one of the performance requirements (constructs). To derive a particular construct score the appropriate items across all assessment instruments were pooled and a mean computed. In this way nine construct scores were computed. Table 3 displays the reliabilities (i.e., coefficient alphas) computed for each of the nine construct scores and the intercorrelations among them.

The majority of our constructs were moderately to highly reliable. We were unable to reliably assess “Management and coordination of multiple battlefield functions”, perhaps due to our operationalization. The five measures derived from the STEP tool were
averaged to produce the DDD score which significantly correlated with eight of the nine constructs.

**Validation**

The nine construct scores and the STEP Composite Score (DDD), were examined separately for Sergeants and Staff Sergeants because of established differences. Demographics and reliabilities for the simulation scores are depicted separately for Sergeants and Staff Sergeants in Table 4.

Correlations among all HumRRO and STEP performance criterion scores are presented in Table 5. The patterns of correlations shown in Table 5 are similar to those reported for the undivided sample in Table 3. That is, construct scores were highly correlated with each other, and less so with the DDD Composite. More interestingly, however, the correlations between the construct scores and the HumRRO criteria were low. The low correlations between the construct scores and HumRRO performance criteria are consistent with past research that has distinguished between “can-do” and “will-do” aspects of performance (e.g., Borman & Motowidlo, 1997; Campbell &

| Table 4. Descriptive Statistics for the Construct Scores, DDD Composite, and HumRRO Criteria by Rank |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|
| Criterion                      | Sergeants       | Staff Sergeants | 
|                                | n   | M      | SD   | ryy  | n   | M      | SD   | ryy  |
|--------------------------------|-----------------|-----------------|-----------------|-----------------|
| STEP                           |     |        |      |      |     |        |      |      |
| DDD                            | 35  | 3.58   | 0.94 | .49  | 30  | 3.44   | 0.73 | .41a |
| Oral                           | 36  | 4.93   | 0.94 | .91  | 31  | 5.19   | 0.93 | .92  |
| Lead                           | 36  | 5.07   | 0.94 | .89  | 31  | 5.01   | 1.06 | .95  |
| Adapt                          | 35  | 4.85   | 1.19 | .84  | 31  | 4.62   | 1.34 | .79  |
| Safe                           | 35  | 4.80   | 1.16 | .68  | 31  | 4.89   | 1.15 | .66  |
| Inter                          | 35  | 4.33   | 1.09 | .74  | 31  | 4.06   | 1.03 | .65  |
| Mgmt                           | 35  | 4.46   | 1.42 | .    | 31  | 4.03   | 1.90 | .    |
| Oral                           | 36  | 5.26   | 0.78 | .81  | 31  | 4.85   | 0.85 | .83  |
| Info                           | 35  | 4.75   | 0.87 | .69  | 31  | 4.59   | 0.89 | .74  |
| Comp                           | 35  | 3.88   | 1.41 | .84  | 29  | 4.02   | 1.50 | .76  |
| HumRRO                         |     |        |      |      |     |        |      |      |
| OPR                            | 41  | 5.05   | 0.92 | .49  | 36  | 5.46   | 0.66 | .41  |
| EFP                            | 41  | 4.80   | 1.40 | .36  | 36  | 4.97   | 1.13 | .80  |

*Note. ryy = Estimated reliability. a One item that averaged into the DDD Score had an item-total correlation of -.21 among Staff Sergeants. The alpha with this item removed was .57. b The Management and Coordination of Multiple Battlefield Functions was based on only one item, thus, no inter-item reliability estimates were generated.*

**Table 5. Correlations among Aptima and NCO21 Performance Criteria by Rank**

<table>
<thead>
<tr>
<th>Criterion</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDD</td>
<td></td>
<td>.10</td>
<td>-.01</td>
<td>.25</td>
<td>.15</td>
<td>-.08</td>
<td>.11</td>
<td>.26</td>
<td>.09</td>
<td>-.13</td>
<td>.06</td>
<td>.21</td>
</tr>
<tr>
<td>Oral</td>
<td></td>
<td>.25</td>
<td></td>
<td>.90</td>
<td>.64</td>
<td>.56</td>
<td>.61</td>
<td>.61</td>
<td>.64</td>
<td>.77</td>
<td>-.15</td>
<td>.04</td>
</tr>
<tr>
<td>Lead</td>
<td></td>
<td>.24</td>
<td>.91</td>
<td></td>
<td>.79</td>
<td>.74</td>
<td>.56</td>
<td>.79</td>
<td>.69</td>
<td>.76</td>
<td>-.19</td>
<td>.00</td>
</tr>
<tr>
<td>Adapt</td>
<td></td>
<td>.30</td>
<td>.71</td>
<td>.83</td>
<td></td>
<td>.68</td>
<td>.42</td>
<td>.76</td>
<td>.78</td>
<td>.62</td>
<td>-.09</td>
<td>.08</td>
</tr>
<tr>
<td>Safe</td>
<td></td>
<td>.34</td>
<td>.63</td>
<td>.64</td>
<td>.54</td>
<td></td>
<td>.30</td>
<td>.74</td>
<td>.48</td>
<td>.54</td>
<td>-.16</td>
<td>-.03</td>
</tr>
<tr>
<td>Inter</td>
<td></td>
<td>.28</td>
<td>.68</td>
<td>.64</td>
<td>.48</td>
<td>.36</td>
<td></td>
<td>.50</td>
<td>.58</td>
<td>.67</td>
<td>-.12</td>
<td>.21</td>
</tr>
<tr>
<td>Mgmt</td>
<td></td>
<td>.32</td>
<td>.60</td>
<td>.68</td>
<td>.65</td>
<td>.56</td>
<td>.52</td>
<td></td>
<td>.72</td>
<td>.66</td>
<td>-.14</td>
<td>-.14</td>
</tr>
<tr>
<td>Prob</td>
<td></td>
<td>.34</td>
<td>.81</td>
<td>.81</td>
<td>.77</td>
<td>.67</td>
<td>.67</td>
<td>.65</td>
<td></td>
<td>.81</td>
<td>-.16</td>
<td>.15</td>
</tr>
<tr>
<td>Info</td>
<td></td>
<td>.26</td>
<td>.75</td>
<td>.72</td>
<td>.59</td>
<td>.58</td>
<td>.60</td>
<td>.50</td>
<td>.82</td>
<td></td>
<td>.02</td>
<td>.32</td>
</tr>
<tr>
<td>Comp</td>
<td></td>
<td>.22</td>
<td>.50</td>
<td>.56</td>
<td>.53</td>
<td>.38</td>
<td>.42</td>
<td>.40</td>
<td>.47</td>
<td>.39</td>
<td></td>
<td>-.14</td>
</tr>
<tr>
<td>OPR (HumRRO)</td>
<td>-.20</td>
<td>.24</td>
<td>.22</td>
<td>.06</td>
<td>.11</td>
<td>.17</td>
<td>.15</td>
<td>.28</td>
<td>.20</td>
<td>.16</td>
<td></td>
<td>.59</td>
</tr>
<tr>
<td>EFP (HumRRO)</td>
<td>-.01</td>
<td>.28</td>
<td>.25</td>
<td>.11</td>
<td>.06</td>
<td>.27</td>
<td>.25</td>
<td>.28</td>
<td>.25</td>
<td>.12</td>
<td>.81</td>
<td></td>
</tr>
</tbody>
</table>
Correlational Results of Predictors and Criteria

Among Sergeants, the SimPPW showed moderate validity for predicting nearly all STEP construct scores, yet little validity for predicting HumRRO criteria. Among Staff Sergeants, the SimPPW showed low, non-significant, levels of validity for predicting both the STEP and HumRRO criteria. For Sergeants, the AIM Leadership score showed moderate validity estimates against 8 of the 10 STEP performance criteria. The AIM Dependability, Agreeableness, and Physical Conditioning scores exhibited some significant negative relationships with the STEP criteria for Sergeants; in contrast, all Sergeant AIM validity estimates were positive in the NCO21 validation sample (Sager et al., 2002). For Staff Sergeants, none of the AIM scores showed significant relationships with the STEP criteria. For Sergeants, the BIQ Social Perceptiveness, Tolerance for Ambiguity, Openness, Leadership, and Social Perceptiveness scores showed consistently positive relationships with both STEP and HumRRO criteria. The BIQ Manipulativeness and Hostility to Authority scores exhibited significant levels of validity, particularly for the HumRRO criteria. For Staff Sergeants, none of the BIQ scores showed consistent relationships with the STEP criteria. Among Sergeants, the ASVAB GT showed low to moderate validity for predicting the STEP criteria, yet little validity for predicting the HumRRO criteria. The only significant Sergeant ASVAB GT validity was for the Interrelationship of Units construct. For Staff Sergeants, the ASVAB GT showed moderately significant levels of validity for the majority of STEP criteria, yet low validity for predicting the HumRRO criteria. Compared to the NCO21 validation sample, the ASVAB GT showed comparable levels of validity for Sergeants, and higher levels of validity for Staff Sergeants on both HumRRO criteria (Sager et al., 2002).

DISCUSSION

The STEP criteria in the present investigation arguably tap the “can-do” aspects of performance to a greater extent than the HumRRO performance ratings do, and thus, as found, were differentially related to the set of set of predictors examined here. Specific differences include the following: the ASVAB GT (a cognitively-oriented...
predictor) tended to be more highly correlated with the STEP criteria than the HumRRO criteria (particularly among Staff Sergeants). AIM and BIQ (temperament-based predictors) tended to be more highly correlated with the HumRRO criteria than the STEP criteria (again, particularly among Staff Sergeants).
Table 6. Estimated Raw Validity Coefficients by Rank

<table>
<thead>
<tr>
<th>Predictor</th>
<th>STEP Performance Criteria</th>
<th>HumRRO Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DDD</td>
<td>Oral</td>
</tr>
<tr>
<td><strong>Sergeants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulated PPW</td>
<td>.10</td>
<td>.48</td>
</tr>
<tr>
<td>SJT</td>
<td>.02</td>
<td>-.10</td>
</tr>
<tr>
<td>Dependability¹</td>
<td>-.25</td>
<td>-.16</td>
</tr>
<tr>
<td>Adjustment¹</td>
<td>.03</td>
<td>.07</td>
</tr>
<tr>
<td>Work Orientation¹</td>
<td>.26</td>
<td>.01</td>
</tr>
<tr>
<td>Leadership¹</td>
<td>.31</td>
<td>.42</td>
</tr>
<tr>
<td>Agreeableness¹</td>
<td>-.04</td>
<td>-.35</td>
</tr>
<tr>
<td>Physical Conditioning³</td>
<td>-.41</td>
<td>-.29</td>
</tr>
<tr>
<td>Hostility to Authority²</td>
<td>-.13</td>
<td>.05</td>
</tr>
<tr>
<td>Manipulativeness²</td>
<td>-.11</td>
<td>-.09</td>
</tr>
<tr>
<td>Social Perceptiveness²</td>
<td>.15</td>
<td>.18</td>
</tr>
<tr>
<td>Social Maturity²</td>
<td>-.01</td>
<td>-.15</td>
</tr>
<tr>
<td>Tolerance for Ambiguity²</td>
<td>.37</td>
<td>.23</td>
</tr>
<tr>
<td>Openness²</td>
<td>.18</td>
<td>.26</td>
</tr>
<tr>
<td>Leadership²</td>
<td>.40</td>
<td>.34</td>
</tr>
<tr>
<td>Interpersonal Skill³</td>
<td>-.10</td>
<td>-.16</td>
</tr>
<tr>
<td>ASVAB GT Score</td>
<td>.28</td>
<td>.13</td>
</tr>
<tr>
<td><strong>Staff Sergeants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulated PPW</td>
<td>.15</td>
<td>.11</td>
</tr>
<tr>
<td>SJT</td>
<td>.00</td>
<td>.27</td>
</tr>
<tr>
<td>Dependability¹</td>
<td>.26</td>
<td>.10</td>
</tr>
<tr>
<td>Adjustment¹</td>
<td>.20</td>
<td>-.10</td>
</tr>
<tr>
<td>Work Orientation¹</td>
<td>-.01</td>
<td>.11</td>
</tr>
<tr>
<td>Leadership¹</td>
<td>.06</td>
<td>.21</td>
</tr>
<tr>
<td>Agreeableness¹</td>
<td>-.03</td>
<td>-.04</td>
</tr>
<tr>
<td>Physical Conditioning³</td>
<td>.08</td>
<td>-.04</td>
</tr>
<tr>
<td>Hostility to Authority²</td>
<td>-.04</td>
<td>.20</td>
</tr>
<tr>
<td>Manipulativeness²</td>
<td>.15</td>
<td>.13</td>
</tr>
<tr>
<td>Social Perceptiveness²</td>
<td>-.05</td>
<td>.09</td>
</tr>
<tr>
<td>Social Maturity²</td>
<td>-.13</td>
<td>-.14</td>
</tr>
<tr>
<td>Tolerance for Ambiguity²</td>
<td>.24</td>
<td>.35</td>
</tr>
<tr>
<td>Openness²</td>
<td>-.01</td>
<td>.05</td>
</tr>
<tr>
<td>Leadership²</td>
<td>.16</td>
<td>.17</td>
</tr>
<tr>
<td>Interpersonal Skill³</td>
<td>.10</td>
<td>.00</td>
</tr>
<tr>
<td>ASVAB GT Score</td>
<td>.21</td>
<td>.53</td>
</tr>
</tbody>
</table>

Note. *N*_{Sergeant} = 33-41. *n*_{Staff Sergeant} = 28-36. All correlations are raw. Statistically significant correlations are bolded, *p* < .05 (one-tailed). ¹ Constructs derived from AIM. ² Constructs derived from BIQ.

These findings are consistent with past research which has shown temperament constructs to be more related to “will-do” aspects of performance, where as “can-do” aspects of performance are more related to cognitive ability (Borman & Motowidlo, 2002).

44th Annual Conference of the International Military Testing Association
Ottawa, Canada, 22-24 October 2002
1997). In conclusion, using simulation methodology shows promise in performance measurement for situations where the goal is to assess the individual’s actual ability to perform the work.

**REFERENCES**


As we enter the 21st century, the military is in the middle of a major transition. Due to changes in mission and technology, the organizational and occupational characteristics that have come to define the present day military are being overhauled. As a result of this process it is imperative to develop a full understanding of the role that enlisted personnel play in the “new military.” This role includes interface with systems, equipment, and personnel that may appear quite a bit different than in the past. What individual requirements will these players need to accomplish their mission? How will this translate to personal readiness? How will performance be defined and measured?

In addition to individual requirements for successful performance, a number of other factors play important roles in the success of selection and classification in the military, as in other organizations. The military uses results from large-scale aptitude testing as its primary basis for making selections into the service. Following this initial selection, testing results are further utilized in making classification decisions. The mechanism for making a personnel classification is a somewhat complicated process that involves using a combination of individual ability, obtained from the aforementioned testing program, the needs of the specific service in question (regarding jobs that need to be filled and gender and minority quota requirements), and individual interest. Historically, interest has received the least amount of weight.

Following classification into a job, an individual will go through basic training, then proceed to the training school pipeline that has been proscribed for the particular
assigned career path. After finishing the initial training pipeline, an individual will be put on the job, complete the first term of enlistment, then reenlist or not. A number of other factors, in addition to the things an individual brings into the service, play a crucial role in how that person perceives their military career and whether the choice to reenlist or not is made. Organizational variables have typically received little or no attention in the military services evaluating reasons for attrition or retention. Additionally, alternate views, owing more to economics or consumer decision-making than to the I/O psychology field, have received less than warranted interest. Currently, Navy researchers are delving into the aforementioned areas and attempting to fill in the knowledge gaps that exist. These efforts are focused toward developing a state-of-the-art personnel system that will ensure the meeting of individual sailor needs, while preventing attrition and encouraging retention.

As mentioned previously, personal interest has taken a backseat to aptitude test results and Navy needs. As an extension of recent Navy developments in the actual classification engine, Navy researchers have developed a computer-administered vocational interest inventory that presents Navy job information as a function of a process and content model. This inventory has attracted a good deal of attention within the Navy’s personnel management community and promises to significantly alter the way Navy jobs are marketed and eventually improve the current status of new enlistee perceptions toward the classification process. David Alderton will discuss the current status of this research program (JOIN – Jobs and Occupational Interests in the Navy) and future developmental thrusts.
Organizational variables, and the role they play in selection and classification are also attracting a significant amount of research interest in the Navy. Long known to be important in employee retention, Navy researchers have undertaken a rigorous program to develop measures of person-organization fit that will prove beneficial in sailor career counseling, and may one day play an important role in the overall classification process. Jackie Mottern will discuss the background and development of “First Watch,” emphasizing the longitudinal nature of the effort to follow sailors from their first day of enlistment to the time they leave service. An extension of this work has focused on the development of a Navy-specific commitment scale and ways to utilize such a measure. Bill Farmer will discuss results from analysis that provide promise for this scale.

Since, selection and classification in the new Navy will require a greater understanding of where the individual fits in the system, significant research is being undertaken to clarify the individual requirements for performance and the criteria upon which these predictors will be evaluated. These efforts are directed toward the development of a future assessment battery that will take into account non-cognitive factors, in an effort to better place future sailors. Jerry Hedge will talk about NAPOMS (Navy Psychometrics of Measures) and discuss some of the implications.

Though much of the promising selection and classification research in the Navy is being done by I/O psychologists, economists are looking at team decision-making utilizing a combination of individual decision theory under conditions of uncertainty and game theory. In efforts that are directed at selecting individuals into team working conditions, this research offers a fresh approach that will enlighten and perhaps improve
future processes. Tanja Blackstone will discuss this research and highlight its implications.
Title: JOIN: Job and Occupational Interest in the Navy

Authors: Paul G. Michael, University of Memphis, Regina Hindelang, University of Memphis, William L. Farmer, Navy Personnel Research Studies and Technology (NPRST), David L. Alderton, Navy Personnel Research Studies and Technology (NPRST), & Stephen E. Watson, Department of the Navy

A new measure of vocational interest, Job and Occupational Interest in the Navy (JOIN), is being developed for use in conjunction with the Rating Identification Engine (RIDE) to help provide a better match between a recruit's abilities, interests and specific occupations (i.e., ratings). JOIN not only informs recruits about enlisted Navy ratings, but measures interest in specific work activities and environments.

There were four objectives in the design of JOIN. First, it had to differentiate among the 79 entry-level Navy jobs, something civilian and other military interest measures could not do. Next, it had to be “model based” so that it could be quickly adapted to Navy job mergers, changes, or additions. Third, JOIN needed to be useable by naïve-enlisted applicants; naïve in terms of knowledge of Navy jobs and the technical terms used to describe them. Finally, JOIN needed to be short and engaging to encourage acceptance by Navy applicants and those who process them.

Model development for JOIN was iterative; with subject matter experts employing abbreviated job analytic procedures to create brief work activity statements called process-content (PC) pairs, which serve as individual interest items (e.g., maintain mechanical equipment). A total of twenty-six PC pairs were judged to cover all enlisted work activities. The goal was to identify process and content words that were generic enough to be
comprehended by naïve applicants while portraying an accurate description of Navy ratings. Along with work activities, other work dimensions relative to Navy jobs (e.g., communities and environments) were included in the model and linked to specific ratings.

The current version of JOIN is computer-administrated; participants indicate their interest in over 300 pictures and text that represent the PC pairs, seven communities (e.g., submarine), work environments (e.g., outdoors) and work styles (e.g., teamwork). Three hundred new recruits in basic training evaluated this version for usability and efficacy.

Overall the results show that JOIN was well liked by the participants. Recruits reported that JOIN was easy to use (89%) and was visually appealing (84%). Thirty-one percent of the participants indicated that they used the pictures to make decisions regarding their interest while 23% reported that they used both text and pictures.

Participants were asked to indicate their level of interest in the work environments, styles, and activities using a 100-point scale; ranging from 0 (“not interested”) to 100 (“very interested”). Importantly for classification requirements, all of the PC pairs, communities, environments, and work styles were endorsed as interesting, and there were substantial and reliable individual differences. More recruits ranked the aviation community as their first choice than any other community area (33%). Participants indicated a preference for physical work (M = 70.2, SD = 28.02) in an outdoor environment M = 79.6, SD = 27.31). Each work activity was presented three times (with different images each time). Recruits were least interested in operating facilities (M = 34.8, SD = 32.03) and most interested in operating weapons (M = 73.9,
SD = 31.24). The (alpha) reliability estimate across all work activities was very good ($\alpha = 0.91$). Reliability estimates for individual work activity items (3 each) were also good, ranging from 0.83 (operate mechanical equipment) to 0.95 (make facilities).

The initial results from the usability testing are very promising on several levels. First, the feedback items provide researchers with an overall positive evaluation of the quality of the computer-administered interest inventory. Second, descriptive statistics of JOIN items indicate substantial variance across individual responses. In other words, the participants were different in their level of interest in various items. Finally, the reliability of the work activity items appear to be very consistent in measuring participant interest in the individual enlisted rating job tasks. Next steps include validation of the SME model and large-scale data collection to explore the relationship between JOIN, selection tests, classification decisions, and training success.
Title: Retention and Attrition Among New Recruits

Authors: Jacqueline A. Mottern, Michael A White, & David L. Alderton, Navy Personnel Research Studies and Technology (NPRST)

Managing retention and attrition is a topic of interest to all organizations and is critical to their success. As a large employer, the U.S. Navy faces problems with recruiting and retaining a quality force of Sailors. Annually, the US Navy enlists nearly 50,000 individuals. Historically, 12.2% (6,100) will not complete basic training and an additional 22.6% (13,300) will not complete their 48-month service contracts, producing nearly a 35% overall attrition rate (17,400) for the first-term of enlistment. When recruits fail to complete their obligation (i.e., attrite), the Navy suffers monetary costs from losing the service member and this cost is compounded by the cost of replacing that individual. Beyond direct costs, there are costs from the turbulence: fleet instability, reduced readiness, lower morale, and excessive burden on remaining personnel. Moreover, the recruit returns home and carries a negative message that may reduce enlistment propensity, elevating the cost of subsequent recruiting.

This paper reports data from a comprehensive investigation of the first-term of enlistment and places retention and attrition in the larger context of organizational and individual differences theory and research. The purpose of this research is to help the Navy better understand the career decisions made during the critical first term of enlistment and to use that information to help
produce high quality sailors who are well prepared for their careers in the Navy and for life.

This research effort will follow an entire one-year cohort of new recruits (approximately 50,000) through their first four years of enlistment. Measures of person-organization fit, commitment and stress coping skills have been developed for this effort and are included in four specially designed questionnaires that are being administered at key points in recruit training. The first questionnaire is designed for new recruits as they begin Navy training and includes questions concerning their recruiting experiences, the classification process through which they are assigned a rate or “job” in the Navy, their reasons for joining and their background. The second questionnaire is being administered eight weeks later at the end of their initial basic training and, in addition to repeating some of the same items on recruiting and classification, adds questions on recruits’ training experiences and the “command climate” of their training division. The third questionnaire is being administered at the end to their advanced (‘A’ school) training course or at the end of their apprentice course (for those recruits not assigned to advanced courses). The apprentice schools last only a few weeks, and the advanced courses vary in length. This questionnaire provides an update on their A/Apprentice school experiences. For those recruits who drop out of training or are separated from the Navy, a fourth questionnaire will be administered prior to separation. This questionnaire asks about their training experiences and reasons for leaving the Navy. After training, new sailors will
be tracked using the web-based Argus Career Milestone Tracking System during the rest of their first term of enlistment. Additional data from recruiting and training databases will be linked to the cohort database. Data collection for the cohort began April 2, 2002 and will end March 31, 2003.

Attrition in the Navy is traditionally defined as failure to complete the first term of obligated service. Rather than look at attrition as a dichotomous variable, we examine attrition rate as a continuous variable measured by weeks of service. We define how quickly a recruit leaves the Navy as “survivability” and theorize that low levels of person organization fit, low commitment, poor stress coping skills, low levels of perceived social support, unmet expectations of the Navy, and negative experiences in training are related to survivability. The Navy Fit Scale, the Navy Commitment Scale and the Navy Recruit Stress Coping Scale are being validated with this cohort of recruits and detailed results will be presented. Data from the command climate sections of the questionnaires indicate recruits change their attitudes toward the Navy as they gain experience in the Navy and with increased knowledge of the Navy thus emphasizing the importance or realistic previews of the Navy during recruiting. The majority of recruits credit their Navy training with improving their self-discipline, self-confidence, and physical fitness and report the Navy as their best current career choice.
Title: Development of the Navy Commitment Scale

Authors: Jacqueline A. Mottern, Michael A White, & William L. Farmer, Navy Personnel Research Studies and Technology (NPRST)

A great deal of attention has been given to the study of commitment to organizations and the link between commitment and organizational turnover. In 1999, the Navy Personnel Research, Studies and Technology Department (PERS-1) began development of a commitment scale appropriate for use with Navy personnel. We developed a model based on a commitment scale developed by Meyer and Allen (1987). The Meyer and Allen model of commitment has three components: affective, continuance, and normative.

NPRST modified the Meyer and Allen scale only to the extent that the word “Navy” was substituted for “organization” and “company.” The items were included in questionnaires on attrition and retention issues that were pre-tested with Navy personnel. An additional set of items reflecting values-similarity were also included and tested. After administration to three samples (approximately 40,000 Marines, 50,000 Navy Reservists, and 35,000 Sailors) at different locations, the data were factor analyzed and compared with self-reports of intention to remain in the service past one’s obligation.

The resulting Navy Commitment Scale is comprised of only two of the original three components: affective and continuance. The normative subscale did not discriminate among Sailors and showed such small variation that it was not useful. Upon closer examination of that component and interviews with sailors, the normative component was based heavily on loyalty to a company or organization. Loyalty, a key value of the Navy, is so deeply entrenched in our Sailors as to prove useless in
differentiation. Further correspondence with the U. S. Army Research Institute indicated a similar finding in their work on commitment (also based on Meyer and Allen). A third component developed by NPRST, values similarity, tested well in the pre-tests and was added.

NPRST believes that the Navy Commitment Scale will prove to be a kind of “leading economic indication” for Navy attrition and re-enlistment. The scale was included in the Argus Career Milestone Decision System, the Navy’s web-based transition survey that was deployed in Feb 01 and the 1st Watch Project currently collecting data on a one-year cohort of recruits through their first term of enlistment (approximately 50,000 new Sailors). This will allow us to test the utility of the Navy Commitment Scale in actually predicting turnover in the critical first term of enlistment.

Preliminary results, based on logistic regression analyses, have shown that the commitment scale does a good job of predicting intentions for those that definitely know that they will remain in the service until retirement. Results also point to the positive predictive ability for those that are inclined toward leaving the service prior to retirement. Results were weaker and mixed for those whose intentions were unclear.
Title: Research Directions for the Future of Navy Selection/Classification


The Navy is in the midst of profound changes that are likely to result in different occupational and organizational structures in the future. To respond effectively to these changes may require operating in a wide and varied set of situations, cultures, and environments. Military personnel will likely need to be more versatile and able to handle a wider variety of diverse and complex tasks. At the very least, for the Navy, each sailor will be required to perform a broader range of tasks, have more sophisticated technological knowledge and skills, and operate more independently with fewer coworkers and a truncated chain of command.

To start a Navy enlisted career today, an individual must take a battery of tests, known as the Armed Forces Vocational Aptitude Battery, which measures basic cognitive ability. However, beyond these cognitive attributes that predict ability to perform effectively in military jobs, individuals possess a variety of preferences, interests, and personal characteristics that should be useful for identifying who will be best suited for military missions of the future. Historically, the preponderance of military predictive validation work has
centered on measuring success in basic and technical training, with job performance in the first-term of enlistment included as a criterion measure sporadically. However, because finding and training a 21st Century sailor will be much more complex and costly than it is today, success on the job beyond the first term of enlistment in the Navy will be increasingly important. This presentation examines the state of the science in both predictor and criterion domain, with an eye toward future Navy personnel selection.

Advances in the last decade or so have shown that we can reliably measure personality, motivational, and interest facets of human behavior and that under certain conditions these can add substantially to our ability to predict attrition, retention, and school and job performance. The reemergence of personality and related volitional constructs as predictors is a positive sign, in that this trend should result in a more complete mapping of the KSAO requirements for jobs and organizations, beyond general cognitive ability. One particularly promising approach to measuring individual differences in the interpersonal and personality areas is the situational judgment test (SJT). These tests are based on the premise that there are important and often subtle differences between the behavior of effective and ineffective persons as they respond to problems or dilemmas confronted in the course of carrying out their job responsibilities and that such differences are reflected in their responses to similar situations presented in written form.

Research has demonstrated that short-term, technical performance criteria, particularly overall school grades, are best predicted by general intelligence while longer term, more differentiated criteria such as non-technical
job performance criteria, retention, and promotion rates are better predicted by other measures, including personality, interest, and motivation instruments. In order to select and retain the best possible applicants, it would seem critical to understand, develop, and evaluate multiple measures of short-and long-term performance, as well as other indicators of organizational effectiveness such as attrition/retention. This presentation explores in greater detail criterion measurement domains such as contextual performance, adaptability, and work commitment.

In general, then, when one considers what attributes are most relevant to perform effectively in any given job, there are many from which to choose. The type of person characteristic viewed as important to success in a job may vary from situation to situation. For example, for a job or set of jobs, one may be most interested in choosing persons that have high cognitive ability, and care much less about their personality or interest patterns. In other situations the reverse may be true. For optimal assignment, it is necessary to somehow link the attributes to how necessary they are for effective performance in specific jobs or job types, and as attempts are made to expand the predictor and criterion space, it will be important to extend one's perspective to broader implementation issues that involve thinking about classification and person-organization (P-O) fit. As organizational flexibility in effectively utilizing employees increasingly becomes an issue (e.g., workers are more often moved from job to job in the organization), the P-O model may be more relevant in comparison with the traditional person-
job match approach. The usefulness of these models will be explored in greater detail in the presentation.

The focus of the current work, then, has been to consider ways of expanding the predictor and criterion domains such that it results in selecting applicants with a greater chance of long-term career success in the Navy.
Title: **Teams, Strategic Games and Optimal Outcomes**

**Authors:** Tanja F. Blackstone, Navy Personnel Research, Studies, and Technology (NPRST), & Peter M. Williams, University of Southern Alabama

The economic theory of the team addresses a middle ground between the theory of individual decision under uncertainty and the theory of games. (Game theory is the study of individual/group strategy that involves a planned sequence of moves and a set of reactions of moves that rivals might make). A team is made up of a number of decision-makers, with common interests and beliefs, but controlling different decision variables and basing their decisions on (possibly) different information. The theory of teams is concerned with (1) the allocation of decision variables (tasks) and information among the members of the team, and (2) the characterization of efficient decision rules, given the allocation of task and information.

A formal model of team theory:

Consider a team with $M$ members. Each member $m$ controls an action $a_m$. The resulting utility (satisfaction) to the team depends on the team action,

$$a = (a_1, \ldots, a_M)$$

and on the state of the environment. (An assumption is imposed on the model that team members have common interests, there is a single utility for the whole team). The state of the environment comprises all the variables about which team members may be uncertain before choosing their actions. If we denote the state of the environment by $s$, ...
then we can denote the utility to the team by $u(a,s)$, (this states that utility or satisfaction is a function of the teams actions and the state of the environment); the function $u$ will be called the payoff function for the team.

Before choosing an action, each team member $m$ receives an information signal, $y_m$. This information is determined by the state of the environment, say $y_m = ?_m(s)$. Let $?_m$ be denoted the information function for member $m$, and the M-tuple $? = (?_1,...,?_M)$ will be called the information structure of the team.

Each team member $m$ will choose his action on the basis of the information signal he receives, according to the decision function, say $?_m$. Thus

$$a_m = ?_m(y_m) = ?_m(?_m(s))$$

Equation (1) states that the action of the team member is dependent on the information set available to the team member given the state of the environment.

Letting $?$ denote the team decision function, then the utility of the team in state $s$ can be expressed as:

$$U(s) = u(?,?_m(s), s)$$

Equation (2) states that the teams utility (satisfaction) is dependent on (1) the teams decisions, which is itself dependent on the information available to the team given the state of the environment, and (2) the state of the environment.

Equation (2) specifies $s$ as known to the team. Alternatively, uncertainty can be introduced into the model by allowing $s$ to be determined according to some probability distribution, $?$, on the set $S$ of possible states. Under the assumption that $s$ is drawn from
a probability distribution, \( \mathcal{P} \), then the team maximizes an expected utility function given as follows:

\[
E[U(s)] = \mathcal{P}(s)U(s)
\]

The expected utility of the team depends on the team members’ decision function, the team information structure and the probability distribution of states of the environment, as well as on the ‘structure’ of the decision problem, i.e. the way in which utility (2) depends on the members’ action and the state of the environment. Thus the optimal problem for the team may be posed in two stages: (1) for a given information structure, characterize the optimal team decision function(s); (2) optimize the information structure, taking account of costs of – or constraints on – making the information available.

**Preliminary Results**

A primary objective of this research was to ascertain in teams can make optimal decisions when faced with complex decisions. The research focused on the quality (optimal choice) and not quantity (productivity) of team choices. Undergraduate subjects were recruited and randomly assigned to two and three person team. Subject groups were informed that cash awards would be given at the end of the experiment session and that cash awards were contingent on performance. Subject groups were asked to search a matrix populated by random numbers with a value range of 100 to 1000. Subject groups were asked to jointly select a combination of cells that maximized their payoff (cash award) subject to a budget constraint. In other words, the team’s reward from selecting \( k \) cells is the sum of the rewards from each of the cells that the team selects.

Subject maximizes

\[
(1a) \quad \max \sum_{i=1}^{k} \text{Cell Payoff} \cdot \text{Cell Weight} \cdot \text{Cell Value} \\
(1b) \quad \max \sum_{i=1}^{k} \text{Cell Payoff} \cdot \text{Cell Weight} \cdot \text{Cell Value} \\
(2) \quad \max \sum_{i=1}^{k} \text{Cell Value} \cdot \text{Value Limit}
\]
Preliminary findings indicate that for small teams, team size does not inhibit the ability to complete the task. Relative to individual subjects, two member teams perform better, (have higher relative earnings). In comparison, three member teams perform worse than their two-member team counterparts. Team size may effect intermittent steps teams adopt to achieve task completion, however, it is yet to be determined if intermittent tasks are an optimal path. Determining optimal path may be critical to enhancing productivity in complete environments. Further, future research controlling for heuristics and time are planned.
INTRODUCTION

Military capability, the ability to deliver combat effect, consists of a number of components. In the Australian Defence Force (ADF), one of the fundamental inputs to capability is personnel. The personnel component of military capability is often considered simply in terms of the size and structure (including skill sets) of the workforce. Few assumptions are made about the individuals in this equation, other than they meet the basic requirements of their job or qualification. This is an objective consideration of personnel capability.

People, however, are inherently subjective and a perspective on personnel capability that is wholly objective will be deficient, there must be consideration of the subjective component of personnel capability. Generally, attempts to define or measure this subjective component of personnel capability have proved difficult. In essence the subjective component of personnel capability represents the quality of the people in the organisation, and measuring this is not easy, defining it. One concept that has recently gained currency in recent times is Well being and this is a concept that may enable consideration of the subjective part of personnel capability.

A Model of Well Being

The concept of well being is broader than the term wellness with which it is often confused. The term wellness can be considered to represent the opposite of illness, and it can be conceptualised as the opposite end, from illness, on a continuum that has the absence of any health concerns as the midpoint.

The term well being is much broader than the simple health connotations of the term wellness. Other issues that impact on the subjective part of capability include issues such as social support, family support, etc. The term well being is therefore much broader and it can be operationalised as shown in Figure 1:

Maximum Ill being                              Just being                 Optimum well being

|____________________________________|________________________________|

Figure 1: A Conceptualisation of the Well Being continuum
Here, the midpoint reflects an individual who is meeting the basic requirements of their life, they are coping with work, relationships, family life, etc. Those to the left of this point are, in fact, not coping with some part of their life (perhaps work or relationships), while those to the right are in fact performing above the minimum standard, and value-adding either to what they do or to their relationships with others.

In the workplace, the mid-point might represent an individual who just meets all of the competencies required in their job statement and their life outside work does not impact on their work performance. Any individual to the left of the mid-point represents a cost to the system because of some deficit (health, emotional stability, etc), which is manifested either directly through medical costs or indirectly through a loss of productivity.

Any individual to the right of the mid-point represents a gain to the organisation. Importantly, any person to the right of the mid-point who experiences some difficulty with their life, might experience some decrement in their well being, but if they are far enough to the right, this might not move them past the mid-point so that they do not become a cost to the organisation. Thus the value of well being programs lies in their ability to improve the robustness of individuals so that they can experience some difficulty, yet still remain productive in their family, social and work life.

This conceptualisation also allows us to consider the effect that the external environment (family and social life) can have on the work environment. It provides the link between the work and non-work environment that may allow policy makers and planners to identify all of the facets that influence personnel capability rather than the less inclusive model that has been used in the past. It also provides the subjective component of capability to complement the objective component of capability.

But what are the elements of well being? There are many definitions of well being, or wellness. Stratton\textsuperscript{42} discussed the importance of defining the term correctly and identifies the satisfaction of important physical needs, the sense of control one has over one's life and the sense of purpose that exists for the individual. She also identifies the importance that social institutions can play in the well being of the individual. Clearly then, well being must incorporate both the physical and mental wellness of the individual as well as make some reference to their social environment. Given this, a proposed definition of well being for use by the ADO is:

\begin{quote}
Well being is a multi-dimensional concept that includes physical and mental wellness, emotional resilience social support, and family support. It represents the individual's ability to be productive and contribute to their work and life. It represents the subjective component of personnel capability.
\end{quote}

OPERATIONALISING WELL BEING: THE ADO WELL BEING FORUM

\textsuperscript{42}Stratton, P. 2002. \textit{Defining Wellness? The impact on the definition on the direction of the ADFMHS}, unpublished service paper, Department of Defence, Canberra, Australia.
There are three key issues in operationalising this definition so that it can contribute to the ADF’s operational capability:

a. determining exactly what are the elements of well being,
b. identifying programs that will contribute to moving the ADF member to the right on the well being continuum, and
c. measuring well being, both in terms of baseline measures and in terms of the impacts of any well being program that are put in place.

In the ADF all of these issues are being addressed through the establishment of the ADO\textsuperscript{43} Well Being Forum. This came about as the result of discussions among a number of areas in the ADO concerning the apparent wide range of functional areas that contribute to the subjective component of capability. It seemed at the time that there might be value in having a forum for communication and collaboration about well being issues within the ADO. The Forum has attracted a wide range of participants from those areas of Defence that perceive themselves as making some contribution to the wellness of ADO members.

**The Elements of Well Being: Forum Participation**

Initially, those involved in establishing the Forum invited anyone that they thought would be interested to participate in the initial meeting. Since then, and by agreement of the participants, participation in the Forum has been left open to any organisation that feels that it has a role in contributing to the well being of ADO members. Forum participants have agreed to monitor the membership of the forum to ensure that it remains relevant and not unwieldy. Other key issues include that it has been agreed by participants that membership of the Forum should be equitable and there is no requirement for a specific lead agency.

Membership of the forum has included representatives of health (including mental health), psychology, social work, the Defence Equity Organisation, single service personnel agencies, the ADO infrastructure organisation, chaplains, workplace relations, and policy makers from the employment conditions area.

**Well Being Programs: The Role of the Forum**

The role of the Forum was discussed at first meeting; and while there was a sense that the Forum could inform strategic personnel policy in the ADF, the general consensus was that the role should initially be coordination and communication.

The forum was seen as having a role in allowing the various elements within the ADO that contribute to ADO personnel well being to share information about what initiatives were on going within the organisation. There was a sense that support to ADO

---

\textsuperscript{43}ADO – Australian Defence Organisation; this term refers to all uniformed and civilian members of the Department of Defence.
personnel could be streamlined and improved the coordination of the many initiatives that were occurring.

It was also felt that was a need to share information about specific issues that may be common to different organisations. There was a sense that there may be issues common to many organisations that, in isolation, are not seen as having the broader implications that might be the case when they are seen across the organisation.

**Measuring the Contribution of Well Being: The Real Challenge**

The real challenge in implementing Well Being in an organisation is being able to measure the impact of any programs that are put in place. An early realisation was that to be successful within a large bureaucracy like the ADO, it would be necessary to show a link to the "bottom line" of Defence. It was also realised, however, that this would present a significant challenge.

It certainly is possible to measure well being in the civilian sector. Research from the Corporate Leadership Council\(^44\) clearly indicate that some organisations actively measure the impact of their programs. This research showed returns on investments in different companies ranging from $220-$300 per employee, and return rates of between three and seven dollars per dollar spent on well being programs. Measuring ROI in the military is somewhat less clear, however a number of strategies have been identified for use in measuring the effect of the Forum.

The Australian Unity Well Being Index is a national measure of Australian's views on a range of economic and social indicators in Australia at both a personal and national level\(^45\). The Index focuses on the following aspects of life:

a. health,
b. standard of living,
c. achievements in life,
d. personal relationships,
e. sense of personal safety,
f. community connectedness, and
g. future security.

These items have been incorporated into the next ADO General Attitude Survey to be administered early in 2003. This will be administered to a 30% random sample of the ADO and allow comparison between the ADO and the general Australian population.

While this approach will provide a good indication of ADO member's perceptions there is still a clear need to link any initiatives of the Forum to observable behaviours.


\(^45\)Taken from the Introduction section of the Australian Unity Well Being Index web site, [www.australianunity.com.au](http://www.australianunity.com.au)
This will allow a link the subjective component of personnel capability to the objective measure of personnel capability in order to produce a complete picture of personnel capability.

As a result, one of the key future tasks of the Forum is to identify how more objective measures of well being, as well as examining what existing research may exist within the ADO that can contribute to this.
CONCLUSION

The concept of well being has a clear and direct application to the ADO in terms of recognising the subjective component of personnel capability and identifying that this can be measured and these measures incorporated into consideration of ADF capability. The definition that has been proposed identifies the key personal, internal, and external contributors to well being, and these have been operationalised through participation in the ADO Well Being Forum.

The challenge for the ADO now lies in identifying appropriate measures of well being that are objective enough to be comparable with existing measures of objective personnel capability and combined and incorporated into ADF capability planning.
Screening for Adjustment Difficulties After Peacekeeper missions

Lieutenant Colonel M.J. Levey  
Commanding Officer, 1st Psychology Unit  
Australian Defence Force, Canberra, Australia  
Martin.Levey@defence.gov.au  

Colonel A.J. Cotton,  
Director of Mental Health  
Australian Defence Force, Canberra, Australia  
Anthony.Cotton@defence.gov.au  

Professor M. Creamer  
Director  
Australian Centre for Post-traumatic Mental Health  
m.creamer@melb.edu.au

INTRODUCTION

The Australian Defence Force (ADF) has very actively used psychologists in direct support to operations for the past fifteen years. There are a number of factors that contributed to this level of involvement. First was an increasing awareness in the Army that operational service can have a significant, negative psychological impact on soldiers. Second was the growth of traumatic stress studies and the development of psychological debriefing as a standard response to traumatic incidents. A third key factor was the increase in operational tempo arising from the increased involvement in peace support operations by the ADF.

Aim

The aim of this paper is to describe the development of the post-deployment support model used by the ADF. The history of the model will be discussed, how the screen differs from the extant debriefing paradigms will be discussed, and finally, the research strategies that have been proposed to evaluate the model will be discussed.

The History of the Development of the Debriefing Model

The main initial task undertaken by psychologists in support to operations was the debriefing of personnel returning from operations, there were many determinants of this, including humane concerns and command concern about the effectiveness and retention

---

This paper draws in large part from a chapter from The Fire Within: The impact of stress on military operations, the Australian Defence Force experience. Melbourne University Press, (in press).
of personnel. Fortunately AA Psych was not without debriefing experience - thanks to a long history of support to the Australian National Antarctic Research Expedition – but most of this experience was limited to a small proportion of the psychologists in the Corps at the time.

Due to their base training and professional experience, however, most AA Psych officers quickly developed the skills necessary for debriefing. These include group facilitation, a sensitivity to indications of serious stress reactions as preventive assessment, a comprehensive understanding of stress and coping, and at least the rudiments of psycho-social care (Raphael, 1986b).

It was apparent that the preferred trauma debriefing model in the civilian sector – Critical Incident Stress Debriefing (CISD: Mitchell, 1983; 1988) – was inappropriate in an operational setting as the majority of service personnel in most missions did not experience traumatic stress. And even if they did, the incident would often have occurred weeks or even months before, making CISD inappropriate due to the extended timeframe. A tailored approach was needed, for example, it was common that organisational stressors often generated as much strain as trauma, see Table 1 below (Murphy, 1990). This result has since been repeated in other research in Australia (Collyer, 1995a; 1995b) and in New Zealand (MacDonald et al., 1998).

**Table 1**

*Percentage of respondents from five Australian contingents who found given stressors caused ‘extreme’ stress*

<table>
<thead>
<tr>
<th>Stressor</th>
<th>ASC Sinai</th>
<th>ASC Western Sahara</th>
<th>ASC Cambodia</th>
<th>ASC Somalia</th>
<th>ASC Rwanda</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threat of danger</td>
<td>0.0</td>
<td>2.4</td>
<td>17.8</td>
<td>22.2</td>
<td>5.2</td>
</tr>
<tr>
<td>Health concerns</td>
<td>9.1</td>
<td>21.4</td>
<td>17.5</td>
<td>6.3</td>
<td>9.4</td>
</tr>
<tr>
<td>Sorting out problems at home</td>
<td>22.3</td>
<td>21.4</td>
<td>26.9</td>
<td>18.8</td>
<td>10.8</td>
</tr>
<tr>
<td>Lack of concern shown by Army</td>
<td>38.9</td>
<td>7.2</td>
<td>31.0</td>
<td>31.3</td>
<td>10.9</td>
</tr>
<tr>
<td>Double standards</td>
<td>31.8</td>
<td>35.7</td>
<td>36.3</td>
<td>18.9</td>
<td>48.0</td>
</tr>
<tr>
<td>Mutinational force and observers</td>
<td>22.7</td>
<td>26.1</td>
<td>49.2</td>
<td>50.0</td>
<td>21.4</td>
</tr>
</tbody>
</table>

Another interesting finding that emerged in many debriefings was that most personnel coped well with traumatic stressors such as being the target of weapons fire. ‘Critical’ incidents tend to be characterised as bizarre, unexpected or involving the suffering of other human beings. There was also some evidence of ‘expedition fever’ (Fiennes, 1983) at work, where minor issues, which would cause minimal concern in a normal environment assume great importance in the isolated operational environment. The resilience that helps military personnel cope seems to be related to both their training and to their self-concept as professionals who are doing what they are trained to do and
who expect unpleasant sights and experiences. Many, however, experience difficulty in coping on their return to the routine and safety of Australia’s society.

Debriefing following operations developed several primary aims. These included providing awareness of operational stress, providing personnel the chance to ventilate about issues relating to their experience, identifying personnel who required more immediate follow up from their deployment; and gaining some systematic data about the human dimension of operational service. Issues of separation and the impending homecoming were also included in debriefings in an effort to prepare personnel for the challenges of reunion and transition. The process also included completion of an anonymous questionnaire on some of the human aspects of deployment. And, finally, individual follow-up, either in-theatre or after return to Australia, occurred when the need was recognised or it was requested.

Debriefings were initially conducted in groups, following the prevailing best practice, however, experience quickly showed that individual debriefs were preferred, particularly if the individual had personal issues to discuss. Experience also showed that debriefings conducted in the Area of Operations were more effective, however, for small operations, debriefs often occurred immediately on return to Australia.

Hand in hand with operational debriefing was the development and conduct of pre-departure briefings. These identified the likely stressors of the deployment and enhanced awareness of pragmatic and effective stress coping skills. Effective pre-briefings were characterised by:

- small group numbers, enabling discussion rather than straight instruction
- use of service personnel who had prior exposure to the mission area as co-presenters
- a focus on a few, specific stressors such as isolation
- pragmatic information on cultural aspects of the mission
- the absence of time pressures

Psychological Support Teams

The need to provide in-country debriefing support meant that there was a need to maintain a deployable psychology capability. The key drivers for this were an understanding that psychological casualties are an inevitable byproduct of war and that the majority of these casualties can be returned to duty within days if basic principles of management are implemented. Although this had previously been seen as a medical or psychiatric responsibility, the time-proven principles of casualty management involved basic techniques that were perhaps better suited to a behavioural as opposed to a medical paradigm.

Thus there was a need for a deployable psychology capability. Over a period of about six years this capability developed from Combat Stress Management Teams into Operational Stress Management Teams, and ultimately, into Psychological Support Teams.
Teams (PsSTs). This was more than a semantic change – the composition of the team changed from a team of four to two personnel, consisting of a psychologist and an NCO. It also reflected a broadening of psychological support activities required in the deployed environment.

In 1998, the Army Land Command, the operational component of the Army, identified, established a structure where PsSTs were incorporated into deployable health units. With this structure in place, emphasis was placed on training and policy as pathways to fostering credibility and acceptance of psychology in operational roles, specific training courses were developed in both operational psychology. Most important, perhaps, has been the refinement of pre-departure briefings on topics such as operational stress, hostage survival, tactical negotiation and cross-cultural competence and awareness. The dissemination of the ADF publication entitled *Operational Stress Management* (Department of Defence, Surgeon General, 1997) gave further impetus and official sanction to the stress management role largely delivered by Defence psychology. The net result of these initiatives and opportunities was that by the end of the decade, most psychology positions within Land Command were firmly integrated as deployable assets. Four PsSTs were deployed to East Timor in 1999 in a number of primary roles. These teams were supplemented by further PsSTs on a short-term basis to conduct return-to-Australia debriefings for elements of Operation Warden in the early months of 2000.

Table 2 shows the range of duties and responsibilities that PsSTs have undertaken in recent years. While not all these tasks are routinely conducted, there is a discernible trend that psychologists and examiners are being called upon to do more. By way of example, over a six-month period in East Timor, PsST activities included 126 self-referrals, 74 medical referrals, 47 administrative referrals, 169 operational or Critical Incident Stress Debriefings, 26 in-service assessments, 249 presentations and more than 5200 return-to-Australia briefings (Michalski, 2000).

**Table 2**

*Overview of the activities of the deployed Psychological Support Team*

<table>
<thead>
<tr>
<th>Duties and responsibilities</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provide critical incident stress management</td>
<td>Routine</td>
</tr>
<tr>
<td>Coordinate and provide Return-to-Australia debriefings</td>
<td>Routine</td>
</tr>
<tr>
<td>Monitor the human dimension of operations</td>
<td>New role</td>
</tr>
<tr>
<td>Provide advice on command and human dimension issues</td>
<td>As requested.</td>
</tr>
<tr>
<td>Provide routine selection services (for example, corps transfer, specialist postings, education courses)</td>
<td>Routine</td>
</tr>
<tr>
<td>Liaise with medical specialists to provide mental health assessments, interventions and referrals</td>
<td>Routine, within professional capabilities</td>
</tr>
<tr>
<td>Conduct research tasks, e.g., trend analyses of issues such as sick-parades, disciplinary offences.</td>
<td>New role; as requested</td>
</tr>
<tr>
<td>Inform commanders of the role and capabilities of Psychological Support Teams</td>
<td>Routine</td>
</tr>
<tr>
<td>Provide advice and support to commanders and contingent</td>
<td>Routine</td>
</tr>
</tbody>
</table>
members on mental health issues.

Provide training, advice and support to commanders and contingent members on operational effectiveness issues, e.g., fatigue management; psych aspects of the deployment cycle, body handling

Liaison with other national contingents on military psychology matters

Provide advice to repatriation cases

| Expanding role. | Irregular | Routine |

As can be seen from the numbers presented above, most personnel see psychology staff only for a return-to-Australia briefing which occurs during the final weeks of their tour. This briefing is the result of previous experience with debriefing and has two components: a group educative session on generic homecoming issues, and individual debriefings focused on personal experiences of the tour. The activity has recently been renamed Return to Australia Psychological Screening.

A subsequent contact with the individual normally occurs between three and six months after deployment, this is referred to as the Post Operational Psychological Screen. A self-report mental health screen is administered at this stage, which includes a measure of serious stress reactions (PCL-C), alcohol use (AUDIT), and general mental health (GHQ12). As well as an individual interview, members are given psycho-educational handouts on topics such as reactions to trauma and the homecoming stage of deployment, and advised of a range of available support options if they encounter difficulties. Personnel identified as having existing adjustment problems are offered support.

A key component of this model is the provision of advice to commanders about mental health and other command issues that exist within the unit. Early debriefings were less comprehensive in the provision of this advice to commanders and suffered from a lack of command support as a result. Current practice, however, enjoys a high degree of command support.

Conclusion

The past decade has seen an expansion in the types and tempo of operations undertaken by the ADF. Allied to this change has been the evolving support provided by military psychology and a gradual recognition of the value of the various contributions by defence psychologists to these operations. Within the Army, pre-deployment briefings, RTAPS, and POPS are now integral to operational deployments.

Defence psychologists have developed a comprehensive psychological screening model that draws heavily on previous experience with debriefing and they also provide highly tailored pre-briefings for operational personnel.
The Australian Defence Force Mental Health Strategy

Colonel A.J. Cotton,
Director of Mental Health
Australian Defence Force, Canberra, Australia
Anthony.Cotton@defence.gov.au

INTRODUCTION

The impact of military service, particularly of combat, on the mental health of soldiers has been known for centuries. Mention is made by writers in antiquity such as Homer and Xenophon to the maladies that afflicted soldiers in their time, and the use of psychiatric diagnoses was common in the American Civil war. Psychiatry was a relatively new field at the outset of the First World War and a major branch of medicine by the end of the war. Since then, the growth of mental health services in the military has been enormous.

This rise in the use of mental health services in the military has strong parallels in the general community also. Quality of life surveys, at least in western societies, routinely show that mental health rates highly, if not highest, among the issues of greatest concern to people. This is supported by mental health prevalence data; current Australian research indicates that in any one year nearly one in five Australians will experience some symptoms of a mental health problem that will significantly interfere with their functioning.

Given that military forces reflect the issues of their parent society, there will be a growth in the need for mental health services in the military irrespective of the operational commitments of the force. The Australian Defence Force (ADF) is no different; the scope of mental health problems facing the ADF and its commanders include:

? Mental disorders are the second leading cause of invalidity retirements from the ADF (ADF Health Status Report 2002, unpublished).
? One in six ADF members misuse alcohol (Sobering Facts; Options for an ADF Alcohol Management Program. Tri-service Working Party Report, Department of Defence, Canberra, April, 2000.).
? Suicide is the second leading cause of death among Royal Australian Navy personnel, the primary cause for Army, and the fourth leading cause of death for Royal Australian Air Force members (ADF Health Status Report 2002, unpublished).

The cost of the defined burden of mental health problems on the ADF is estimated to be around $20m per annum (ADF Health Status Report 2002, unpublished). The undefined burden, being the impact on people other than those directly affected, is more difficult to measure, but is characterised by ongoing family problems for the member, reduced job performance (including discipline and morale problems in the member’s unit), and possible separation from the ADF. The hidden cost of mental health problems is...
defined as the increase in mental health problems that occur as the result of individuals not seeking adequate or early support for their problems due to the stigma attached to mental health problems, again, calculating this for the ADF is very difficult.

Because of the direct effect that it has on the individual and their circumstances, mental health is a key determinant of the personnel component of military capability and as such it has a direct influence on military capability. The scope of mental health impacts on capability is very complex and the need for a comprehensive Mental Health Strategy for the ADF was recognised in the ADF Health Status Report 2000 (Department of Defence).

Due to the complexity of the mental health issues facing the ADF, the delivery of mental health services to the ADF must be based on a set of principles that allow the application of services in a broad range of situations. This approach is consistent with recent changes in ADF policy making where there has been a recognised move away from rules-based to principles-based policy. Development of the Strategy had to take into account the approach taken to mental health within the general community, the existing mental health services in the ADF, and the key drivers for the delivery of mental health services in the ADF.

The Australian Mental Health Environment

The Australian National Mental Health Strategy was developed in 1992 (Reference B). This Strategy emphasises the need for a comprehensive and integrated approach to mental health. While this strategy focussed on specific client groups for mental health, none of which relate exactly to the ADF population, there were a number of key features of the National Strategy that informed the development of the ADF Mental Health Strategy (ADF MHS) delivery of mental health services to the ADF; these include:

? The distinction between diagnosable mental disorders and the much wider category of mental health problems.
? The need for increased emphasis in prevention and early intervention of mental health problems; including the development of better strategies for promoting mental health.
? The need for an increase in research and evaluation studies in the mental health area and for the development of strategies for monitoring the outcomes of mental health standards.

The Australian Department of Veteran’s Affairs has also developed a comprehensive mental health policy that is also aligned with the National Mental Health Strategy (Australian Government 1992). The ADF and DVA have a very strong relationship in the area of mental health services through the a number of initiatives, including the DVA Links project (Department of Veteran’s Affairs, 2000).

Mental Health Services in the ADF

The ADF Health Status Report 2000 recognised that mental health plays a key
role in an individual's overall health and that mental ill health may have an effect on a member's ability to function at an optimal level. It also identified that mental health services and policy development in the ADF are distributed across a number of agencies including: the Defence Health Service (DHS), the Defence Force Psychology Organisation (DFPO), Defence chaplains, and the Defence Community Organisation (DCO). All of these agencies provide a range of services that have not been formally coordinated at an ADF level previously.

There is a broad range of extant ADF mental health policy including: entry standards, critical incident stress management, operational stress management, preventative health, and mental health surveillance. Among the recommendations in HSR 2000 was a need for a comprehensive ADF Mental Health Strategy to review these policies to ensure that they are comprehensive, coordinated and reflect current best practice.

Key Drivers in the Delivery of Mental Health services to the ADF

Given the National mental health environment, and the then current status of mental health service provision in the ADF, four key considerations in the provision of mental health services to the ADF were identified; they were:

- that the mental health of ADF personnel is a key to the personnel component of ADF capability, and the provision of operational capability is the primary reason for providing mental health services to a military force.
- that the ADF screens individuals prior to entry and therefore, effectively starts with a healthy, in a mental health sense, population.
- that the provision of mental health services is a multi-disciplinary activity.
- that military mental health, as a developing area, has a critical need for information, for both surveillance and research purposes.

Mental Health and Operational Capability

Mental health is a state of successful performance of mental function, resulting in productive activities, fulfilling relationships with other people, and the ability to adapt to change and to cope with adversity. Mental ill health has been recognised as the result of either diagnosable mental disorders or mental health problems (National Mental Health Strategy, Government of Australia, Canberra, 1999):

1. Mental disorders are diagnosable health conditions that are characterised by alterations in thinking, mood, or behaviour (or some combination thereof) associated with distress and/or impaired functioning.
2. Mental health problems represent a range of more common mental complaints and include reactions to life stressors that are not necessarily diagnosable (within a normally accepted diagnostic framework), but that affect peoples' lives and functioning none-the-less.
This distinction between diagnosable medical conditions and the broader lifestyle issues that can none-the-less have a deleterious effect on individuals is important in terms of providing a framework for understanding the breadth of issues that will require services to maintain or improve mental health in the ADF. This had not been adequately addressed previously in the ADF, so the first principle for the delivery of mental health to the ADF is:

**Principle 1: Mental ill health can be the result of either a diagnosable mental disorder or a mental health problem.**

This allows the ADF to clearly identify that, irrespective of the cause, mental ill health will diminish the sufferer’s ability to function, both at work and at home; and it will therefore diminish an individual’s ability to contribute to ADF capability. Secondly, it allows us to recognise that, diagnosable mental disorders are medical problems that require treatment from within the medical system. But that the broader range of mental health problems, on the other hand, need not necessarily be treated in the medical system, rather there are a range of intervention choices for mental health problems.

It also provides a clear link to the responsibility of the commander by identifying that because mental ill health will cause a diminution of personnel capability, the individual will need to be managed by the command chain, because the maintenance of the personnel component of capability is a command responsibility. Therefore, the management of the mental health of ADF members is a command responsibility, and the provision of mental health services to the ADF is a support activity for commanders in the execution of their responsibilities.

**Principle 2: The management of mental health in the ADF is a command responsibility and any mental health services delivered to the ADF must support commanders in the exercise of their responsibilities.**

**A Healthy Population**

One of the key features of the ADF is that it screens individuals prior to entry and therefore, it can be assumed that the ADF population is healthier, in a mental health sense, than the general population. In order to maintain its capability then, the ADF needs to maintain the mental health of its members, this is an outcome of mental health promotion, which aims to optimise mental health and well being in communities and thereby in individuals. It focuses on improving environments (social, physical and economic) that effect mental health and enhancing the coping capacities of communities as well as individuals.

**Principle 3: The ADF MHS will emphasise mental health promotion and the prevention of mental disorders or mental health problems.**

Despite the best aims of promotion and prevention, mental health problems and mental disorders will occur in ADF members; these might be the result of their service, or
they may be the result of life cycle issues. Either way, in order to maintain operational capability, the ADF must be able to offer treatment options to its members, and there is a range of effective evidence-based treatments available for both mental disorders and mental health problems. Because of its impact on operational capability the ADF should provide appropriate treatment and management of any mental health problem in ADF members.

**Principle 4:** ADF members should be able to receive appropriate evidence-based treatment and management for any mental health issue irrespective of its cause.

Mental disorders and mental health problems will occur both in an area of operations (AO) and out of the AO in the National Support Area, and ADF commanders expect to be able to receive mental health services for their personnel in both environments. Again, this principle supports the notion that operational capability is primary.

**Principle 5:** Mental health services are to be available to commanders in both an AO and the NSA.

Another major consideration is the role that the family plays in the mental health of the member. The member's family and social support are a significant part of the individual's psychosocial environment and can have a major impact on a member's mental health and therefore on the member's effectiveness. Therefore, consideration of family issues must be included in the provision of ADF mental health services.

**Principle 6:** The provision of mental health services to the ADF must include consideration of family issues.

Finally, given that the ADF starts with a healthy population, there is an obligation on the organisation to ensure that when a member leaves the ADF they are as well as when they joined, or treatment is provided for anything that has occurred as a result of their service. And any separation from the ADF, especially those occurring on administrative or disciplinary grounds, may have a mental health component. Similarly, in many cases, the mental health impacts of service are not seen until some time after the member discharges. There should be scope for the provision of mental health support to these individuals. Those who discharge on grounds of (mental) invalidity will require more intensive management than those separating for other reasons. The immediate and extended provision of mental health support to any member who leaves the ADF is a vital element in the management of separations from the ADF.

The DVA Mental Health Policy recognises the need to provide a seamless transition from military service to civilian life. The key success factors for transition include effective coordination between ADF and DVA, early intervention and rehabilitation as appropriate, flexibility, portability, resource efficiency and national consistency, and ensuring that appropriate on-going levels of care are available for a
member's management. The DVA/Defence Links project exists to provide this and needs to continue to be actively supported by the ADF as part of the provision of adequate mental health services to ADF members.

Principle 7: Access to on-going mental health support for service-related mental health problems should be available to all members who discharge from the ADF.

Multi-disciplinary mental health services

The National Mental Health Strategy's recognition of the separation of mental health problems and mental disorders is consistent with the way that ADF commanders view mental health. A concomitant of this is that the provision of services to support mental health in the ADF is not the sole domain of any one area, but should be the result of the cooperation of a range of providers. There is a clear need for the establishment of a multi-disciplinary framework to deliver and manage mental health services in the ADF; this is consistent with the National Strategy. The current key providers of mental health services to the ADF are: DHSB, DFPO and single service psychology organisations, DCO, and Service Chaplains the should all be included.

Within this broad group of potential mental health treatment providers, levels of expertise will vary. In order to support a multi-disciplinary case management model, there will be a need to provide training, on-going professional development and supervision of ADF mental health providers (doctors and nurses, psychologists, social workers, and chaplains in some cases). There will be a concomitant need to provide a system of credentialling for these personnel, as well as any external providers, and a need to ensure that core mental health providers have an understanding of the specialist military knowledge and philosophy that impacts on the mental well-being of military personnel.

Principle 8: All mental health service providers to the ADF are to be appropriately trained, supervised, and credentialled.

Research, data collection and analysis

The mental health services provided to the ADF need to comply with current and future standards in mental health, both on a national and international level. In order to achieve this, ADF mental health interventions must be evidence-based wherever possible, and have a strong empirical research base to support them. While the research will be conducted by many different agencies both within and outside the ADF, there is a need for coordination and prioritisation of this research. The ADF has recently established a central committee to oversee research in the area of human sciences titled the Health and Human Performance Research Committee (HHPRC). Thisody is capable of providing this coordination of mental health research in the ADF.

To properly evaluate the efficacy of ADF mental health services, there is a need
for a robust mental health surveillance system. This system must be accessible to all mental health providers and must incorporate input from all providers. Issues such as professional confidentiality must be balanced against the broader benefits to the organisation. HealthKEYS will provide the backbone for this, however, there is a need to ensure that the information captured by this system reflects the needs of all potential mental health providers.

\textit{Principle 9: The ADF must conduct, or sponsor, mental health research and surveillance that is aimed at providing evidenced-based mental health services that are properly evaluated.}

Conclusion

Mental health problems cost the ADF a significant amount each year in both measurable and immeasurable ways, it leads to a direct loss of personnel through invalidity discharges, an indirect cost in other discharges that may be reduced, and a significant indirect cost through diminished effectiveness of other personnel. Mental health is a key factor in the personnel component of ADF capability. The provision of mental health in the ADF is currently uncoordinated and inconsistent. If this situation is allowed to remain it there will be an ongoing loss of capability for the ADF.

The ADF currently provides a wide range of mental health services to its members, it needs now to integrate these services, their management, and the policy development activities that support them, in order to secure improvements in the mental health of its personnel. This is consistent with the DHSB emphasis on strategic planning and the development of leadership structures within the ADF to reduce preventable illness. To prepare and provide psychologically robust personnel to commanders and assist them to maintain this capability in the military environment, the principles of delivery of mental health services to the ADF described in this Instruction will be adopted.
DEVELOPMENT OF NEW NAVY PERFORMANCE APPRAISAL AND COUNSELING TOOLS

Walter C. Borman
CEO, Personnel Decisions Research Institutes
100 South Ashley Drive, Suite 375
Tampa, FL 33602
wally.borman@pdri.com
Jerry W. Hedge
Personnel Decisions Research Institutes
Mark J. Bourne
Navy Personnel Command
Michael J. Schwerin
Navy Personnel Command
U. Christan Kubisiak, Kenneth T. Bruskiewicz
Personnel Decisions Research Institutes

INTRODUCTION

The aim of virtually every organization is a productive workforce. One of the most challenging strategic actions, therefore, is the development of employees during their tenure with the organization. The effective performance of individuals is a key to a productive, high quality workforce. Accurate evaluation, performance feedback, and supportive counseling are crucial components of any performance appraisal and development process. This paper highlights the impetus and developmental activities undertaken to create a new performance appraisal and counseling tool for supervisory and non-supervisory personnel in the United States Navy.

Background

In October of 2000, an "Executive Review of Navy Training" was chartered by the Chief of Naval Operations (CNO) to examine the current state of training in the Navy, to report on their findings, and to recommend actions that would lead to a 'revolution in training.' One result of this executive review was the formation of a Task Force for Excellence through Commitment to Education and Learning (EXCEL). Task Force EXCEL's goal is to revolutionize and revitalize Navy training to provide Sailors the opportunity to succeed and prosper in their professional and personal lives. At the heart of this initiative is what is known as "The Sailor Continuum." The Sailor Continuum is the foundation around which the Navy will identify the knowledge, skills and abilities (KSAs) that Sailors need to possess to be successful. As such, it is the vehicle that will drive all training, education and proficiency requirements for every enlisted rate and officer community.

Once established, the Sailor Continuum will provide clearly defined career paths and milestones, giving Sailors the tools and opportunities they need to grow

The views expressed in this paper are those of the authors and do not necessarily represent the views of the U. S. Navy.
professionally and personally. For each Navy occupational field a continuum will be created, allowing Sailors to see exactly what skills and abilities they must possess, and the corresponding training available, at any point in their career. Thus, Task Force EXCEL identifies new ways the Navy can train, grow, place, and utilize personnel that maximizes their ability to accomplish the Navy mission and make for a more productive and satisfying workplace. One vector on this continuum is called the Performance vector, and this is the focus of the work described in this paper.

OBJECTIVES

The primary objective of this research and development effort undertaken for the Navy, under the auspices of the Navy Personnel Command, Bureau of Naval Personnel, was to develop a performance appraisal system for Navy supervisory and non-supervisory personnel. In addition, a counseling tool was needed to support this appraisal system that would provide supervisors with a mechanism for offering performance feedback, and provide management with a process for offering long-term career counseling. Finally, it was deemed essential that the project adopt as a fundamental tenet, Fleet involvement in all phases of the development process. Because of page constraints, development of the Navy supervisor appraisal system will be described in this paper. Readers interested in a more detailed description of the work are referred to Hedge, Borman, & Bruskiewicz (2002).

APPROACH

In order to meet these stated objectives, several critical issues had to be addressed. First, we had to identify a procedure for capturing the important components of the performance domains for both Navy supervisors and non-supervisory personnel. This procedure had to allow us to gather data from all relevant Navy constituencies (e.g., sea-based, shore-based; surface, air; West Coast, East Coast), consolidate these inputs in such a way as to be representative of all of these constituencies, and do so in a highly efficient manner.

Second, once a preliminary set of performance dimensions had been identified that subject matter experts (SMEs) had verified as providing adequate coverage of the performance domains for both supervisors and non-supervisory personnel, an important next step was translation of these performance dimensions into behavior-based tools that could be used by supervisors to evaluate the performance of personnel under their command, regardless of their occupation or duty assignment, and that would be seen by supervisors and managers as relevant and useful for evaluating performance.

Third, in addition to evaluation of employee performance, it was important to also produce a system that could serve as a mechanism for performance feedback and career counseling. Thus, some system (or systems) must result from this process that would serve both an administrative and feedback/counseling purpose. Our solution to each of these issues is described next.
Defining the Performance Domain

An important first step in developing a framework for coverage of the performance space for both Navy supervisory and non-supervisory personnel was to identify all of the variables that characterize these domains, and determine how they are related. Once this was completed, then it was necessary to reduce the universe of dimensions to a manageable number that would still represent the performance domain.

We chose to follow a process recommended by Borman and Brush (1993), and subsequently used by Borman, Ackerman, and Kubisiak (1994) in a large-scale study with the Department of Labor. Borman and colleagues applied a three-stage process that led to a final definition of their performance domains. The first stage relied on subject matter experts (SMEs) to generate performance statements that they believed captured the performance domain in question. In the current study, we held workshops with representative groups of Navy personnel, asked attendees to envision their ideal performance appraisal system, and then had them name and define the dimensions that represent that system.

Performance Statement Generation

The performance statements to be used in the formation of the performance categories for Navy supervisors were collected in workshops conducted at Naval Station San Diego and Naval Station Norfolk. A total of 55 officers participated. The participants ranged from pay grades O2 to O6, plus W3 to W4. Participants were assigned to aviation, surface force, submarine force, and shore-based activities, with the largest group (47%) being assigned to shore-based activities.

During the workshops, participants were briefed on the nature of the Task Force Excel project, as well as the purpose of the current workshop. Participants were instructed to generate examples of behaviors that represented everyday performance for a wide variety of Navy jobs.

The participants produced a total of 531 performance statements, which were then reviewed and edited by PDRI project personnel, to reduce redundancy, and ensure that behaviors did not represent more than one dimension. Statements that were multidimensional were separated into their constituent components to make them unidimensional. After review, there were 126 supervisor performance statements remaining to be used in the sorting workshops.
Sorting Task

The second stage required that another group of SMEs categorize all these performance statements into clusters based on content similarity. These workshops were conducted with 23 officers at Naval Support Activity Mid South at Millington. The participants were primarily between the pay grades of O3 to O5, and all had supervisory responsibilities.

As with previous workshops, the nature of the project and specific task were described to participants. For this workshop, we had prepared cards that each contained a unique performance statement (thus 126 cards). Workshop participants were asked to sort these cards into categories based on content similarity, and then to label and define each category within the performance domain they had created.

Pooling of Performance Statements and Principal Components Analysis

The final stage relied on data analytic techniques to compare and pool performance statements sorted in Stage Two based on similarity, and then provide a final performance category solution by means of factor analysis. First, we created a similarity matrix by pooling performance similarity information across participants. This was accomplished using the method referred to previously that required computing for each pair of performance statements the proportion of workshop SMEs who sorted both statements in the pair into the same category. This was followed by also computing an indirect similarity index, which indicated for any performance statements pair the degree of correspondence between each of these two statements’ patterns of similarity with all other performance statements. These two steps in the third stage produced a similarity correlation matrix, which was then submitted to a principal components analysis, with a varimax rotation, resulting in the following eight performance categories: (1) Communication Skills; (2) Coaching/Mentoring; (3) Displaying Integrity & Professionalism; (4) Embracing Personal & Professional Development; (5) Leading Change; (6) Leading People; (7) Resource Stewardship; (8) Organizational Savvy.

Developing a Behavior-based Measurement System

Once the performance domain for Navy supervisors had been defined, the next step was to transform these performance categories that cover the domain into a set of performance appraisal tools. While a variety of approaches was possible, we felt it was important to produce a behavior-based system so that the Navy could base both their evaluation and their counseling on concrete behaviors. Consequently, we chose to adopt the critical incident methodology. Originally developed by Flanagan (1954), the critical incident technique is a method for obtaining specific, behaviorally focused descriptions of work or other activities.

A good critical incident has four characteristics: it is specific, focuses on observable behaviors exhibited on the job, describes the context in which the behavior occurred, and indicates the consequences of the behavior. Thus, a good critical incident
describes actual behavior in a specific situation with no mention of traits and no judgmental inferences.
Critical Incident Generation

In our work with the Navy, we applied the critical incident methodology by first conducting workshops with SMEs to generate high, mid-range, and low performance behavioral examples for each performance category in the domain. The supervisor critical incident generation workshop was conducted with 21 officers at Naval Station Mayport. Workshop participants were mostly between the pay grades of O2 to O4. Approximately half of them were assigned to shore-based activities and the remaining participants were assigned to either aviation or surface force activities.

In these workshops, officers were asked to describe examples of performance that represented the performance categories described above. For each performance example, each participant was asked to describe the circumstances leading up to the performance example, what the individual did (or did not do) that made the example effective, ineffective, or average in performance, and finally what happened as a result of the individual’s actions. They were also instructed to indicate which performance category the example represents and its effectiveness level (high, medium, or low). The critical incidents were reviewed by project personnel for clarity and edited accordingly. The officer personnel produced 161 critical incidents and on average, each participant wrote 7.7 critical incidents.

Rating Scale Development

Once sufficient numbers of critical incidents were generated for all categories and levels, we worked with these critical incidents to craft summary statements that described representative behaviors at each of these three performance levels for all performance categories. In doing so, we followed the example of Borman (1979) who devised a Behavior Summary Scale (BSS) rating scale performance appraisal format.

Anchor Retranslation

Once we produced scales for each category, they were reviewed by U.S. Navy project personnel and then evaluated via a retranslation process. To accomplish this step, we again held workshops with Navy SMEs, and asked each participant to sort cards with each unlabeled summary statement for all dimensions into proper performance categories and proper performance levels. Participants were asked to read through each of the summary statements comprising the behavioral anchors, place them into a relevant performance category, and then rate the effectiveness of the behavior described in each performance statement as high, mid-range, or low effectiveness.

The supervisory rating scale retranslation workshop was conducted with 11 officers at Naval Station San Diego. The participants had been on active duty for an average of 16 years, and in supervisory roles for an average of 14 years. They represented pay grades O3 to O6, and all of them had supervisory responsibilities.
Overall, there was a high level of agreement of ratings and category placement for the officers. In 99.2 percent of the cases, the officers sorted the performance statements into the intended category and effectiveness level. This not only allowed us to gauge the preciseness and distinctiveness of summary statements, but also to fine-tune these statements where necessary. The end-result of this process was, then, behavior-based performance appraisal systems for Naval supervisory and non-supervisory personnel.

Developing a Tool for Performance Feedback and Counseling

Finally, once a set of performance appraisal tools was developed for performance evaluation purposes, it was still necessary to devise a system that could serve as a mechanism for performance feedback and career counseling. To accomplish this, we first returned to the 126 behavioral statements that served as the raw material for the sorting task and subsequent development of the behavior summary scales. We reviewed all behavioral statements for each dimension, and extracted additional themes. These themes were used, along with the content included in the behavior summary scale anchors, to provide a more detailed list of behaviors for the counseling system. The rationale here was that we wanted supervisors conducting counseling sessions to have as rich a behavioral depiction of each dimension’s domain as possible to provide maximally relevant feedback to counselees.

Although page constraints did not allow description of the development process for the non-supervisory system, the same process described for the supervisor system was used, and resulted in a nine-dimension performance appraisal system. Both the supervisory and non-supervisory appraisal tools are in the process of being implemented Service-wide.

SUMMARY

We used a methodology Borman et al. (1994) employed previously in developing performance dimensions intended to be relevant for all non-management jobs in the U.S. economy, to derive a dimension set targeted toward supervisors in all communities. An 8-dimension system was derived based on input from 110 officers from virtually all major communities. The behaviorally anchored rating scales appear widely relevant, and the performance evaluation system that employs the scales is nearing implementation. A performance counseling protocol was also developed to supplement the performance evaluation system.

REFERENCES


If You can't get No Satisfaction:  
A study into early drop-out during military training in the Royal Netherlands Army

C.P.H.W. van de Ven & F. Matser

Introduction
Each year the Royal Netherlands Army, or RNLA, needs to take on approximately 3,500 personnel for the complement of privates/corporals with a fixed-term contract. To achieve this quota, a much larger group of interested people is needed in order to have enough candidates who maintain that interest during the application phase and to succeed in getting through the (demanding) physical and mental selection process. At the time that the candidate has completed the phase and is deemed trainable, he or she is of great value to the RNLA. This is due, on the one hand, to the investment already made in the potential fixed-term contractor and, on the other, to the need to fulfil the procurement requirement. The latter aspect is a particularly important factor in times of a competitive labour market.

From the moment that the candidate is taken on, he or she attends a training course at one of the four school battalions of the Training Centre for Basic Training: Central, Airmobile, North or South. Following the three-month General Military Training, which is a few weeks longer at the Airmobile school battalion, the fixed-term contractors attend one or more functional training courses at the relevant training centres. Throughout the process, but chiefly during the General Military Training, personnel drop out of training. As holding onto personnel is increasingly the Achilles heel of the RNLA, the interim drop-out phenomenon is considered a problem. This article devotes attention to the motives of fixed-term contractors for leaving training early, the increase in the drop-out rate during training, measures taken to reduce this drop-out rate and evaluation of these measures.

Article structure

The article consists broadly of two parts: the theoretical part and the empirical part. The theoretical part comprises a description of research conducted in the past into the drop-out rate, resulting in a theoretical model. This theoretical model is then used as the basis for the empirical part. Here, the method of research is first looked at, followed by a description of the research results over the past three years.

Part I: The drop-out rate in a theoretical perspective
For the last few years, a great deal of attention has been devoted to the drop-out rate among personnel. The emphasis in most studies is on personnel voluntarily leaving the organisation, whereby the employee takes the initiative to leave; this is in contrast to

* C.P.H.W. van de Ven (drs.) works as a researcher at the Behavioural Sciences Division/Directorate of Personnel and Organisation/RNLA. Colonel F. Matser (drs.) is head of the Behavioural Sciences Division/Directorate of Personnel and Organisation/RNLA
personnel being obliged to leave the organisation, whereby the initiative lies with the employer.
In the research under discussion, we are faced with a specific form of the drop-out rate, namely that which occurs shortly after entry into the organisation, during training. The employee has therefore not yet become acquainted with his or her eventual job or function. The theoretical description will therefore largely concentrate on this form of the drop-out phenomenon.

**Drop-out models**

A model which was advanced in 1973 by Porters and Steers assumes that employees have specific expectations on entering the organisation. If the organisation and the function do not meet initial expectations, dissatisfaction may arise, increasing the risk of drop out. Other researchers declare that dissatisfaction with work initially leads to the employee developing ideas about possibly leaving the organisation. These ideas lead in turn to estimating the expected value of looking for another job and the cost incurred by resignation. If these estimates are favourable, the employee will look for another job. In addition to the cost of resignation, what is to be gained by resignation and what is to be gained by staying on may also play a part in the estimate.

Any consideration of staying or leaving always includes the economy as a major factor. In times of high unemployment and few vacancies, fewer people will resign from jobs, even if they are dissatisfied. Under more favourable conditions, if there are many jobs available, dissatisfied employees will show a greater tendency to leave than satisfied employees. The relationship between satisfaction and the drop-out rate is therefore stronger in times of low unemployment than in times of relatively high unemployment. An additional theory is that there are different determining factors for dropping out in times of high unemployment than in times of plentiful employment. This means that, depending on the state of the economy, in this case the situation on the labour market, the weight applied to the determining factors may vary.

**Expectations**

A meta-analysis of 31 studies relating to the non-realisation of expectations resulted in the following findings: non-realised expectations have a strong influence on job satisfaction, involvement in the organisation, the intention to leave and the actual act of leaving the organisation. Incidentally, this only applies to expectations concerning aspects of the job which are deemed to be important. In the case of non-realised expectations of irrelevant aspects, the reaction is surprise rather than disappointment.

**Important aspects in a job**

In her study into “Important aspects in a job”, Van Zijderveld put the following question to young people in 1999: What are the three most important things you (would) look for when choosing a job?

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Important aspects in a job</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 3</td>
<td>Man</td>
</tr>
<tr>
<td>1. Salary</td>
<td>88%</td>
</tr>
</tbody>
</table>
Salary is spontaneously named most frequently, and more often by boys than girls. Girls seem to name this less frequently as they get older. For boys age makes no difference. People with a technical education name salary relatively often in contrast to those with a care/healthcare education.

Atmosphere (pleasant colleagues and working atmosphere) is named more often by girls than by boys. Interesting work (variation, interesting, fun and a challenge) is named about as often by both sexes. Young respondents name this more frequently than older respondents.

Van Zijderveld also reports on several other studies in which young people have been asked what they find important in a job. An initial measurement held in 1988 among students in their final year of further education highlighted the following job aspects, in order of importance.

1. a good income (in first place for 75% of the young people; boys seem to be more strongly interested in income than girls);
2. communication in the workplace;
3. involvement in content.

A study of the social environment of young people was conducted by means of interviews with 500 young people in 1988. The young people formed a random selection of Dutch young people between the ages of 12-21. The following ranking was drawn up on the basis of importance:

1. the nature of the work is pleasant, varied and fascinating;
2. good contacts between colleagues, good cooperation, pleasant atmosphere at work;
3. a good salary, earning money yourself, being financially independent;
4. a job must benefit well-being;
5. opportunities to develop;
6. bearing responsibility;
7. good working conditions.

Girls seem to have mainly intrinsic job expectations. These are expectations which relate to the nature and content of the job (such as variety in the work, good contacts with colleagues, independence and being appreciated). Boys have mainly extrinsic job expectations. These relate to job characteristics which serve as a means of achieving a goal outside work (such as a good salary, working conditions and having a career). Their own income is deemed important by more boys than girls.

With respect to what young people think important in their jobs, social milieu forms the basis for a clear difference:

a. Young people from higher social milieus are more development and career-oriented than those from lower social milieus. The higher the social milieu, the more frequently young people emphasise communication, involvement, management, status and career.

b. Middle-class young people think that having a career is more important than those from a higher class, and they want to climb the social ladder via their careers.
c. Young people from the lower class are more materiel-oriented and place a good income at the top of their list.

Young people who start work at a young age (these are often lower class) often have a fairly instrumental attitude to work, while those who start later are more interested in content. Important issues concerning the future, such as choosing a partner, may also influence what people think important in a job.

In a longitudinal study held in 1998 in the school battalions, Roepers and Duel investigated which characteristics of a job are seen as the most important. The table below gives these aspects and the point at which the fixed-term contractors were questioned.
Table 2  Most important aspects of a job at different points

<table>
<thead>
<tr>
<th>week 1 (first day)</th>
<th>week 5</th>
<th>week 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. teamwork</td>
<td>1. salary</td>
<td>1. salary</td>
</tr>
<tr>
<td>2. camaraderie</td>
<td>2. camaraderie</td>
<td>2. camaraderie</td>
</tr>
<tr>
<td>3/ 4. gaining work experience</td>
<td>3. variety</td>
<td>3/ 4. teamwork</td>
</tr>
<tr>
<td>3/ 4. being able to become an indefinite contractor</td>
<td>4/ 5. getting the job you want</td>
<td>3/ 4. work experience</td>
</tr>
<tr>
<td>5. pushing back limits</td>
<td>4/ 5. teamwork</td>
<td>5. getting the job you want</td>
</tr>
</tbody>
</table>

The importance that one attaches to specific job characteristics is clearly related to the point, or the state of progress, at which they are questioned in the training process. In the first week, aspects related chiefly to the working atmosphere are found the most important (teamwork and camaraderie). A notable absence in the top 5 at this point is salary. From a certain point, which cannot be specified more closely using these data, the importance of salary increases (enormously) and, for the fixed-term contractors in training, this becomes the most important aspect of their job. Camaraderie is deemed the second most important at all three measurement points.

**Psychological contract**

In a study previously carried out into the drop-out phenomenon among the school battalion South, in which salary is often cited as a reason for leaving, the importance of the salary is linked to the employees’ psychological contract.

The psychological contract comprises the unspoken (often unconscious) expectations with respect to mutual obligations held by the organisation and the individual. The individual commits to making certain sacrifices and, in return, obliges the organisation to guarantee a good salary, social security and security on the duration of the work.

In psychological contracts a distinction is made between transactional contracts and relational contracts. Transactional contracts are chiefly economic and extrinsically oriented and are specified in time and time-related. This type of psychological contract is mainly found among employees who are taken on by the organisation for a specified period with a view to performing special tasks, for which specific expertise and skills are required. These employees are prepared to work hard in exchange for significant financial rewards. Relational contracts are both economically and non-economically-oriented, but otherwise have a more socio-emotional and intrinsic nature and are open-ended (in the case of contracts for an indefinite period).

It can therefore be assumed that young people in particular who are in service for a short period will develop a more transactional contract. An important expectation of this category of employees is that they are awarded reasonable financial compensation for their efforts. If this does not happen, they will either (try to) terminate the working
relationship or reduce their effort to the level at which the employee believes the contract is once more properly balanced.

Two thirds of Dutch young people indeed estimate that fixed-term contractors earn a good salary. This has been proved by studies into the interest of Dutch young people in a job, conducted in the context of the Defence Labour Market Monitor 1999. About a quarter of the respondents had no clear picture of this and only a very small percentage thinks that fixed-term contractors have no good salary.

**Orientation**

Earlier research into the drop-out phenomenon in the school battalions shows that in particular the first week, and to a lesser extent the second week, is prone to the drop-out phenomenon. The initial period, and in particular the first week, therefore require additional attention.

The first week of entry into an organisation is known as the orientation period in the literature on the subject. This period is treated as a separate entity as this week goes hand-in-hand with a high degree of stress. The large number of role changes experienced by the individual causes this stress. Changes which are known to possess a specific stress value and could perhaps occur in new recruits include: the first full-time job or new job, geographical move, earning an income or an increase or drop in income, serious limitation of social life and separation from the parental environment. Furthermore, stress occurs due to the non-realisation of expectations and as a result of worries that one is incapable of meeting the requirements of the organisation.

**Individual factors**

A number of demographic and person-related variables may play a part in people leaving the organisation. Firstly, research has shown that the drop-out rate is higher among younger employees than among older employees. This relationship may partly be explained by other variables. A younger employee generally has more opportunities and fewer responsibilities than an older employee.

There is also a negative link between period of service and leaving the organisation. The drop-out rate is relatively higher in the early years of service than in later years.

Demographic variables such as sex, education and civil status may also have an effect. No systematic links have been demonstrated between personality traits, interests, intelligence and capacities on the one hand and the drop-out phenomenon on the other. This has also been proved by previous research by the Behavioural Sciences Division.

**Information**

The adaptation of newcomers during the initial period depends partly on the precision and completeness of the information given to the individual before entering service. Incomplete and imprecise information can lead to non-realisation of expectations and a reality shock. The extent to which the newcomer has realistic, precise and full information at the time of entering service results in positive attitudes and a lower drop-out rate.
Job versus organisation

Earlier research into the drop-out phenomenon highlighted the fact that some of those who decide to leave the RNLA do so on the basis of not obtaining the function they wanted.

The choice of a specific job in a specific organisation is the result of a series of decisions made during the period of growth into adulthood: this is known as the exclusion process. During childhood, all or nearly all professions are a possibility for the ultimate choice. During puberty, this choice is reduced to the choice for a general professional sector, and in the young adult stage the choice of a specific profession is made. During adulthood, this choice is refined further into the choice of a specific job and the related organisation. The essence is that people vary with respect to the image they have of their work. The most general image relates to a desired professional sector (military). A further refinement applies if people have already opted for a specific profession (military driver). Finally, people may already have made a choice of a specific job and the related organisation (military truck driver).

Research model

The principle of the theoretical description is that the extent to which the initial expectations can be met determines how satisfied people will be with their new job and ultimately may determine their decision to leave the organisation. This only applies to the expectations concerning job characteristics which are deemed important.

Salary is still viewed as the most important or one of the most important characteristics. The importance of salary increases the more people’s profiles meet the characteristics given below:
1. young people who come on the labour market at an early age (often the lowest qualified);
2. male;
3. young people from a lower social milieu;
4. people with a contract for a specific period.

In the case of a job as a fixed-term contractor, the fact that Dutch young people expect fixed-term contractors to earn a good salary is also added. This is a positive aspect for recruitment, but this may increase the drop-out rate in view of expectations on this aspect possibly being too high and therefore being impossible to realise.

Other aspects which young people in general and fixed-term contractors in particular deem important are a good working atmosphere (teamwork and camaraderie) and pleasant, interesting work (getting the function you want).

On the basis of the ranking according to importance of job characteristics (table 2), it can be expected that at the start of training reasons for leaving will chiefly be named which relate to the working atmosphere. Although disappointing experiences with respect to camaraderie are a possible reason to leave throughout training, (possible) disappointment about salary will have an even greater impact, with the exception of the period at the start of the appointment.
**Figure 1** Research model for the drop-out phenomenon (van de Ven, 2000)

**Explanation of model (from right to left)**

People do not leave training or the organisation simply for one reason or for clear reasons. People who leave training or the RNLA due to a reason which lies outside their sphere of influence or outside the RNLA do not want to leave the organisation due to dissatisfaction (in the model: satisfaction). Yet for these people, too, something has happened which is not in line with their expectations (in the model: realised expectations). In addition to the various reasons which may be the cause of resignation, the reason for leaving may also be (strongly) related to the point in time of resignation. During the orientation period, other reasons may also be decisive than those which apply in the period which follows.

Only non-realised expectations relating to important aspects of the job (salary, working atmosphere and nature of the job) will lead to dissatisfaction and possibly ultimately to leaving the organisation.

A (more) competitive labour market will result in an increased tendency for people to leave the organisation. This means that static job satisfaction in times of an increasingly competitive labour market leads to a higher drop-out rate.

The extent to which expectations can be realised goes hand-in-hand with how realistic these are. Expectations which are too high are difficult to realise and therefore lead to
disappointment more quickly. Expectations can be aroused by recruitment and information, correspondingly unrealistic reasons to apply for other jobs, and individual factors.

**PART II:**

**THE EMPIRICAL STUDY**

*Study method*
What are the motives for fixed-term contractors to terminate their training early? Can the increase in the drop-out rate be explained? What are the areas for attention to reduce the drop-out rate (in the short term)?

In order to be able to answer this question, the study model given earlier was developed on the basis of the theoretical description. The basic idea of the model is:

> **The drop-out rate, during the initial period of appointment, is determined by the extent to which initial expectations concerning relevant job aspects are realised.**

This applies both if the initiative comes from the individual and if the organisation dismisses the person. Both parties, in principle, have a positive view of each other prior to the act of appointment. If the individual subsequently takes the initiative to leave, then there is something or several aspects which appear not to have gone as he or she had expected, otherwise the person would have stayed on. If the organisation dismisses a person, the organisation therefore experiences a discrepancy between the requirements and the skills of the person, i.e. the person does not meet expectations.

According to the above, there are just as many reasons for leaving as there are relevant job characteristics, or combinations of job characteristics. In order to generate usable data, it is therefore important to discover whether groups of individuals can be distinguished which experience comparable discrepancies.

*The instrument for data collection*

Using the study model and the job characteristics earmarked as relevant, a questionnaire was drawn up in which these aspects have been included. These exit questionnaires are given to all fixed-term contractors who leave training early by employees of the Total Quality Management Office of the school battalions. The response percentage is therefore 100%.

*Those who continue*

In addition to those who leave training, a representative random selection is made from those who successfully complete training and they are also asked to complete the questionnaire. This structure was chosen in order to be able to relate and put into perspective the answers of those leaving training. This structure and the continuous nature of the study mean that the data are also highly suited to evaluating training (continuously).
Timescale

The timescale in which the data for the first (initial) study were collected comprises the period from March 1999 up to and including February 2000. In total, the data from 601 completed questionnaires by those dropping out of training and 389 of those continuing were analysed.

Study results

Fixed-term contractors did not appear to leave training early due to one reason or for clear reasons. In most cases, it was a combination of reasons, in which different motives played a part for different people. On average, a student had approximately reasons for leaving training early. The most frequently cited in 1999 were ‘military life does not suit me’ (46% of all those who left named this as one of their reasons), ‘homesickness’ (28%), ‘circumstances at home’ (24%) and ‘disappointing salary’ (24%).

A few clear differences could be distinguished between men and women. Women more often cited: ‘medical reasons’ (27% versus 14%) and ‘physical load’ (29% versus 12%). In contrast, men more frequently cited ‘disappointing salary’ (26% versus 7%).

The fact that fixed-term contractors had several reasons for leaving training created the need to investigate whether there were common combinations of reasons for leaving, in order to provide insight and in the interest of practicality. The cluster analysis showed that those leaving training could indeed be classified into five categories or clusters. The following reasons for leaving were not included in the analysis: ‘different function within the RNLA’, ‘dismissed’, ‘temporary stop’ and the reason ‘other’. For the first three reasons, this was not a (definitive) loss for the organisation. The reason ‘other’ is too diverse to use as a variable in a statistical analysis.

<table>
<thead>
<tr>
<th>Cluster 1 (n = 127)</th>
<th>Cluster 2 (n = 122)</th>
<th>Cluster 3 (n = 88)</th>
<th>Cluster 4 (n = 59)</th>
<th>Cluster 5 (n = 90)</th>
</tr>
</thead>
<tbody>
<tr>
<td>homesickness</td>
<td>Military life does not suit</td>
<td>Circumstances at home</td>
<td>wrong function</td>
<td>disappointing salary</td>
</tr>
<tr>
<td>possible mission abroad</td>
<td>Medical reason</td>
<td>Homesickness</td>
<td>few career opportunities</td>
<td>military life does not suit</td>
</tr>
<tr>
<td>military life does not suit</td>
<td>Physical load</td>
<td>possible mission abroad</td>
<td>starting studies</td>
<td>better job</td>
</tr>
<tr>
<td>posting to Germany</td>
<td>better job</td>
<td></td>
<td>few indefinite-term contract opportunities</td>
<td>few career opportunities</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>starting studies</td>
<td></td>
</tr>
</tbody>
</table>

By using these clusters we were able to discover which discrepancies were shared by several people. This made it possible to put specific measures in place for certain groups of people. The results indeed showed that each cluster, besides shared combinations of reasons, had a certain profile with regard to personal characteristics, times of leaving and (dis)satisfaction regarding certain aspects of the job.

In the following paragraphs the clusters will not be described in further detail, but the relationship between the clusters and the realisation of expectations with regard to important aspects of the job will be discussed.
**Important job aspects**
The most important job aspect about which the target group had expectations was a **good salary**. This was, however, also the aspect about which there was the greatest disappointment. Of those leaving training and those who continued, almost half thought expectations concerning the working conditions did not come true, or scarcely came true. Women were, incidentally, significantly more positive about the extent to which this expectation had been fulfilled (26% versus 49%).

With respect to appreciation of the salary, those who had completed training were more negative than those who had left training early (4.6 versus 5.6). Account needs to be taken here of the fact that about half of those who left did so before the first salary had been paid.

To the question of where fixed-term contractors expected to earn more, almost threequarters said outside the RNLA, only 5% thought that they could earn more within the RNLA.

With regard to a **good income**, it was clear that all those in cluster 5 (*disappointing salary*) were highly disappointed and, as it were, forced out of the organisation. Disappointing salary was in this case a *push factor*. In view of the low appreciation of the entire group (those who left and those who continued), but mainly the fact that a majority thought it could earn more outside the RNLA, the more attractive salaries elsewhere acted as a *pull factor*. It must not be forgotten here that it was precisely salary which was the most important form of reward for this group.

The second most important job aspect was a **pleasant atmosphere in the workplace**. Here there was a difference of opinion between those who left and those who continued. One in three of those who left early indicated that expectations of atmosphere did not come true, or not fully, compared to 14% of those who continued. The appreciation awarded was unsatisfactory to a relatively large number of people. Incidentally, the other aspects relating to the working atmosphere, such as colleagues and immediate superiors, were awarded high scores by both those who left and those who continued.

Non-realised expectations concerning the **working atmosphere** were an important factor in leaving training for those in cluster 1 in particular (*homesickness/possible mission abroad*). Many experienced the strict military life as such a huge shock on entering service that they decided to leave as early as in the first week, or in the second week. The phenomenon of homesickness can generally be sub-divided into two factors: difficulty in being away from the old, familiar (often parental) environment and difficulty in adapting to a new environment. In view of the low appreciation during this period and the non-realisation of expectations surrounding the atmosphere, the difficulty in adapting to the new situation was more crucial here. Furthermore, the more pleasant the old situation was and the more unpleasant the new situation was for these people, the more difficult it was to make the transition.

As stated earlier, a distinction can be made for cluster 1 between those with homesickness and *no* combination with a possible mission abroad and those *with* the combination of the two. Of those who only named homesickness as a reason for leaving, 67% left within two weeks. They did not succeed in adapting. Those with homesickness in combination with a possible mission abroad succeeded to some extent in adapting to the new environment; they left at a later point in training. The
anticipatory idea of being sent on a mission abroad to a (possibly) stressful environment, whereby their separation from the familiar environment was irrevocable, made them decide to leave training and the RNLA.

The third job aspect which was very important to fixed-term contractors was the content side of the work: it should be fun, varied and interesting. Although the training period was different from the combat-ready period, it was important here to investigate the extent to which there was possible dissatisfaction among the study population. Here a distinction was made between work content and work relationship.

With respect to the work relationship, there were two main aspects which required improvements. The first concerned the respect and appreciation awarded. A considerable percentage of the exit group indicated an insufficient degree of respect and appreciation. Secondly, a large number believed that there were too few opportunities for showing initiative (and more broadly also for taking decisions themselves and not be ordered around all the time). The most important aspect, being able to obtain assistance from NCOs/officers, was appreciated positively.

With respect to the important aspects relating to work content, almost everyone agreed that this was at least sufficient.

There was a strong relationship between job content and whether people were given the function they opted for. Among those who left training early, the percentage which indicated that it did not obtain the desired function was twice as high as among those who continued.

For cluster 4 (not getting desired function) the motive for resigning was clearly taken from disappointment about (future) job content. Military life and the RNLA as an employer appealed greatly, but the future function they were allocated did not correspond with their expectations and was sometimes even viewed as a broken promise.

For cluster 2 (military/physical), too, there was a discrepancy between the expected job content and the reality. This was, however, different from that in cluster 4 with respect to the future function. Those in cluster 2 had difficulty with the current job content. Aspects of training were either too difficult or the pace was too high.

Other work-related aspects
Those in cluster 3 (circumstances at home) also had a number of actual expectations which did not come true or a number of incidents occurred which were not in line with expectations. In most cases, however, these were beyond the sphere of influence of the RNLA or of the person involved. These often concerned deaths, divorces, illness in the immediate family or the partner not agreeing with the choice of a job as fixed-term contractor, with the corresponding possibility of being sent on a mission abroad for a long period. All these reasons were directly or indirectly related to the RNLA, but in particular to the characteristics of a fixed-term contract job: a 24-hour a day job which is sometimes difficult to combine with a family life.

Training
Although the general assessment of General Military Training and Functional Training was satisfactory, there were a number of aspects which may have aided the drop-out rate. For instance, more than a quarter of those leaving training indicated that they found the
pace of General Military Training too high, compared to 5% of those who continued. The study and evening workload was experienced as considerably higher by those who left than by those successfully completed training. With respect to the physical load during General Military Training, those who dropped out indicated more frequently that the aspects in question were ‘too difficult’, or ‘too fast’. The greatest difference concerned ‘field service/exercises’: 24% of those who left training found this too difficult compared to 3% of those who continued. The acceleration in training was viewed as ‘too fast’ by 21% of those who left compared to 10% among those who continued.

Students were confronted with a busy schedule from the beginning. Although it is clear that it would be desirable for several reasons to start more slowly, especially in the early stages, in view of the content of General Military Training and the ultimate goal, there was simply not enough time for this.

The consequences of this strict schedule were that an increased number of people left due to physical complaints (including many female fixed-term contractors) and too high a mental load. In addition, this busy schedule offered no room for acclimatisation (reality shock) to the new living and working environment. In particular during the first week (known as the orientation phase in the literature on the subject) activities should be aimed at assisting the newcomer in conquering the stressful new situation. But after the first week, too, the lack of time meant that there were too few times of rest and free evenings in which a pleasant working atmosphere could develop.

**Increase in the drop-out rate**

If the labour market is becoming increasingly competitive, static job satisfaction leads to a higher drop-out rate. This also applied to the RNLA, even if nothing had changed in comparison to 1998, and an increase could be explained by the large(r) number of vacancies elsewhere. It was clear that the economy is a difficult factor to influence. In order to even neutralise the drop-out rate due to this reason, job satisfaction needed to be increased.

**Areas for attention**

The non-realisation of these expectations and the resulting reasons for leaving were based on a number of bottlenecks. Removing these bottlenecks can raise (job) satisfaction and thus reduce the drop-out rate among fixed-term contractors in training.

The areas for attention were: **working conditions**, these led to dissatisfaction and could not compete with salaries in civilian society; **function allocation**, this should have been carried out more in conjunction with the fixed-term contractors themselves, and **training**, the ultimate goals and the time reserved to achieve them did not correspond well. The pressure of time which was thus created led to a number of undesirable developments, including: a reality shock at the start of training, too high and fast an increase in the physical and mental load and too few rest periods to allow a pleasant working atmosphere to develop.

**STUDY CONDUCTED IN 2001**

On the basis of the above and other study findings, the RNLA has taken a number of measures, and others would be implemented some time later. The following covers in
more detail the measures taken to reduce drop-out rates and the effects that they have had, as well as the remaining reasons behind premature departure.

**Measures taken in 2000**

**Terms of employment**

From 1 June 2000, all 17 and 18-year-old soldiers were granted a monthly pay rise of approximately f150 to f225 guilders (68 Euro to 102 Euro). That applied to both new recruits and soldiers who had already signed their contracts. The starting monthly salary for a 17 or 18-year-old army soldier rose to between f1600 and f1670 (727 Euro and 759 Euro).

**General Military Training**

Based on the problem areas identified that are related to General Military Training, the Training Centre for General Military Training (the OCIO) has made improvements that are most visible in the creation of the Introduction Period. The Introduction Period is intended to reduce drop-out rates during training by:

- allowing trainees to more gradually become accustomed to the transition from civilian life to RNLA culture, and
- meeting the trainees’ expectations as much as possible.

To achieve this, seven main points have been identified for attention during the Introduction Period:

1. reduction of work and time pressures;
2. meeting expectations;
3. promoting the group process;
4. balanced (reduced) information load;
5. involvement of the home environment;
6. structure and implementation of Day 1; and
7. attention for the individual.

**Specialist trades assignment**

It has become apparent that not getting the desired trade assignment is sufficient reason for some trainees to leave the training and the RNLA prematurely. This was especially so in the case of recent school-leavers and especially school-leavers with a particular trades-orientation. As a result of that determination, trades assignment has been taking place in even closer consultation with the trainee since April 2000.

**Physical pressures**

The high physical demands and the resulting injuries were often named as reasons for departure, especially for female trainees. To deal with this, in addition to a more gradual physical build-up thanks to the Introduction Period, trainees who score unsatisfactorily low during an interim measurement are transferred to a Remedial Platoon. This allows trainees to recover from an injury and/or to develop physically to the required level.
Instructors

A final improvement measure introduced in 2000 relates to the quantitative and qualitative staffing of instructor positions. To achieve a more active instructor-trainee interaction, staffing levels of training units have been increased. In order to be able to meet this requirement, instructor positions have been made more attractive by including them in the career-path, making a stint as an instructor a pre-requisite for later assignments.

The position of Training Supervisor has been restored to its original function. Among the Training Supervisor’s tasks are supervising instructors and providing help in situations requiring complex didactic capabilities. Training Supervisors have since been returned to their original function at all training centres.
Measures evaluation

Employment conditions

The proportion of the training drop-outs who indicated that few or none of the employment conditions expectations were met has declined from half to one-third. For trades trainees the percentage has declined from 39% to 27%. The result for salary has risen by one point to where it can (barely) be considered acceptable. Closer analysis, however, shows that how the salary is perceived depends strongly on one’s age. Seventeen and eighteen-year-olds have a better opinion of their salary than their 19, 20 and 21-year-old colleagues, who rate it as unsatisfactory. Trainees who are 22-years-old and older (a very small portion of the trainee population), however, rate the salary as high as their youngest colleagues do. The evaluations of the other employment conditions are unchanged.

In order to determine the degree to which the RNLA can compete on salary with other jobs outside the RNLA, respondents were asked to indicate whether they think they could earn more. For training drop-outs from before August, the percentage of those who thought they could was 74%. After August that figure drops to 68%. For trades trainees, the percentages are 70% and 62% respectively.

This, too, is strongly dependent on age. Of the drop-outs, the percentage for 17 and 18-year-olds has dropped sharply from 74% to 60%, while it has risen sharply from approximately 74% to 90% for the 19, 20 and 21-year-olds.

Of the trades trainees, a sharp decline has been seen among the 17-year-olds (from 84% to 52%) and a sharp increase among the 19-year-olds (from 54% to 80%).

Respondent personal characteristics

In regard to the personal characteristics of the respondents, a noticeable drop has occurred in the average age, and a related increase in the proportion of 17 and 18-year-olds. For General Military Training drop-outs, the average age has declined from 19.0 to 18.7, and the percentage of 17 and 18-year olds has risen from 47% to 60%. For trades trainees, the average age has dropped from 19.8 to 18.8 and the percentage of 17 and 18-year-olds has risen from 42% to 59%. The change among the trades trainees, however, is partly the result of having them fill in the questionnaire earlier, at the end of General Military Training rather than at the end of trades training. When correction is made for this change, the age reduction is nearly four months for the exit population and eight months for trades trainees. That brings the age for General Military Training drop-outs to 18.7 and for trades trainees drop-outs to 19.1.

Inasmuch as there is a minimum age requirement of 17.5 years for military personnel, a decline of eight months for an average age of 19.8 is significant. The maximum decline that is still possible (everyone 18.0 years old at the end of General Military Training) at that age is 13 months.
With the reduction of the average age, and the correlated increased share of 17 and 18-year-olds, the RNLA also has another population within its ranks. Earlier research into drop-out (GW 99-29) had already showed that the younger trainees have more difficulty coping with working under time pressure. Working under time pressure is an important part of military life and the training for serviceman. An increase in the proportion of younger recruits should therefore lead one to conclude that the number of respondents giving “military life doesn’t suit” as their reason for leaving should rise proportionately. Another characteristic of younger trainees is that they are more likely to have feelings of homesickness (GW 01-13). Both characteristics have been corroborated in this study, given the lower average age in these clusters. Changes in the composition of the trainees with a fixed-term contract can, therefore, obscure some of the effects of the measures, such as the effects of the introduction programme, for example.

Effects in regard to aspects of training/trades characteristics

Despite the possibility that some of the effects may be obscured, the measures that have been taken by the OCIO have been sorted by their most important aspects.

Work atmosphere and labour relations, for example, have been rated (somewhat) higher and in regard to aspects of the training, positive developments have been noted in tempo, study load, physical stress and training structure. Some aspects are still being rated more heavily by training drop-outs than by trades trainees, and therefore, continue to demand attention. Trades assignment is being increasingly done in consultation with the trainees, resulting in a reduction in the number of trainees indicating that they have not been assigned the trade that they wanted.

Effects in relation to reasons for dropping out

The reduction or elimination of a problem area subsequently leads to a reduction in the related reasons for dropping out. A comparison was made between 1999 and 2000 for those taking early release because they did not get the trade assignment they wanted. The drop-out percentage over the comparison period declined from 19% to 11%.

Measures related to the other drop-out categories were introduced in (or just prior to) August 2000, making a comparison between the first and second half of that year more meaningful. The drop-out rate in 2000 was 30%. When one looks closer at the recruiting groups from before the introduction of the measures and afterwards, then one sees a difference, with the drop-out rate for pre-measure-period groups at 33% and for post-measure-period groups at 27% (a decline of 18%).

Inasmuch as a reduction in drop-outs occurred during this period, comparing the frequency with which a particular reason was given with the actual in-flow during the same period will give a more representative result.
The results show that four out of five of the most-often named reasons for leaving early have been reduced significantly. Only in the category “Military life unsuitable” is no significant reduction apparent. The motivations “homesickness”, “Salary level”, “Home situation”, and “Possibility of deployment abroad” show a decline of between 28% and 34%.

**STUDY CONDUCTED IN 2002**

**Steps implemented in 2001**

In addition to the measures already discussed, a number of measures were decided in 2000 and implemented in 2001.

**Working conditions:** As of 1 January 2001, starting salaries for new recruits were improved and made more flexible. The base level of compensation for new recruits becomes the current salary of 19-year-old. To prevent individuals salary levels from falling below the legal minimum wage, new recruits who are 20-years-old or older will be moved up by at least one and by no more than four increments. As of 1 October 2001, a salary increase of 8.2% spread over the following two years for military personnel came into effect. The addition of an additional rank meant that the salary increase for soldiers could amount to as much as 15%.

**Training extension:** as of August 2001, the length of General Military Training was increased by a month, and now lasts four months. The reason for this was the fact that more and more problems were occurring in this area. Extra subjects had been added to the training over the years without the duration of the training having been increased. This increased the workload considerably, for both students and instructors, making the training more superficial. The extension of the training period allows, for instance, a more modular approach, which increases clarity, a more gradual increase of the physical load and the reduction of the number of hours from 51 to 42.

**Realistic video:** As of January 2001, the trainees receive a realistic information video prior to beginning General Military Training, but after they were taken on.

In 2001, the RNLA succeeded in recruiting more military personnel for the first time in four years. A steadily downward line seems to have turned upwards for the first time. The

---

**Table 4. Drop-out motivation in relation to actual in-flow, 1st and 2nd half 2000**

<table>
<thead>
<tr>
<th>Drop-out motivation</th>
<th>% of in-flow (1st half: 2nd half)</th>
<th>% difference (2nd half from 1st half)</th>
<th>% decline result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Military life unsuitable</td>
<td>14.9% : 14.0%</td>
<td>0.9%</td>
<td>Reduction by 6%</td>
</tr>
<tr>
<td>Homesickness</td>
<td>9.9% : 6.5%</td>
<td>3.4%</td>
<td>Reduction by 34%</td>
</tr>
<tr>
<td>Salary level</td>
<td>8.3% : 5.8%</td>
<td>2.5%</td>
<td>Reduction by 30%</td>
</tr>
<tr>
<td>Home situation</td>
<td>7.8% : 5.6%</td>
<td>2.2%</td>
<td>Reduction by 28%</td>
</tr>
<tr>
<td>Possibility of deployment abroad</td>
<td>7.0% : 4.8%</td>
<td>2.2%</td>
<td>Reduction by 31%</td>
</tr>
</tbody>
</table>

---
number of personnel entering service with both fixed and indefinite contracts, increased from 3091 (2000) to 3410 (2001), an increase of more than 10%.

Measures evaluation

Terms of employment
The terms of employment, in particular the salaries, were clearly rated higher in 2001. All age categories now give the salaries a satisfactory grade and there is no longer any difference between training drop-outs and those who continue. However, there is still a significant percentage who think that they would be able to earn more outside the RNLA, although this percentage has dropped by 10% to 58% of drop-outs and 49% of those who continue.

Extension of the training
Study load, evening workload and physical stress during field service are all rated as considerably less by both drop-outs and those who continue training. The tempo of the General Military Training is now also more often considered to be appropriate by both groups.

A quarter of the drop-outs, however, had difficulties with the evening workload. Combined with the fact that one in three considers the extension of the training to have resulted in it becoming too long, this is an issue that must be given attention. They have to be able to deal with the workload for an extra month.

The extension of the General Military Training has, however, had no effect whatsoever on the rating of the training as a whole (both those who leave and those who progress to trades training are satisfied). It is remarkable that the extension has had no effect on the extent to which trainees consider the training to prepare them for their jobs in active service. The extended set-up could have been expected to have a positive effect in this respect.

Information video
The introduction of the information video had no effect on experiences versus expectations. This is partly due to the time at which the candidates are given the video, i.e. after they have applied for a job. The initial expectations that they have before they apply (which are their reasons for applying) are therefore not changed by the video. The majority of the trainees are of the opinion, however, that it is an accurate representation of reality. This means that the video can have a positive effect on the ‘vaccination value’, as it immunises the candidates to a certain extent against the coming difficult periods of the training.

Reduction of the reasons for dropping out and the drop-out percentage
After the measures taken in 2001 the drop-out percentage fell further by 15% (from 27% to 23%).
Table 5. Drop-out motivation in relation to actual in-flow, 1\textsuperscript{st} and 2\textsuperscript{nd} half 2001

<table>
<thead>
<tr>
<th>DROP-OUT motivation</th>
<th>% of in-flow (1\textsuperscript{st} half: 2\textsuperscript{nd} half)</th>
<th>% difference (2\textsuperscript{nd} half from 1\textsuperscript{st} half)</th>
<th>% decline result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Military life unsuitable</td>
<td>12.0% : 10.0%</td>
<td>2.0%</td>
<td>Reduction by 17%</td>
</tr>
<tr>
<td>Homesickness</td>
<td>7.4% : 5.9%</td>
<td>1.5%</td>
<td>Reduction by 20%</td>
</tr>
<tr>
<td>Salary level</td>
<td>2.7% : 2.3%</td>
<td>0.4%</td>
<td>Reduction by 15%</td>
</tr>
<tr>
<td>Home situation</td>
<td>5.6% : 5.5%</td>
<td>0.1%</td>
<td>Reduction by    -</td>
</tr>
<tr>
<td>Possibility of deployment abroad</td>
<td>4.5% : 4.1%</td>
<td>0.4%</td>
<td>Reduction by 9%</td>
</tr>
</tbody>
</table>

The reduction in the drop-out percentage is a result of fewer people having reason to leave the training due to 'military life does not suit me' and 'salary level'. The reduction of the reason 'homesickness' is flattering. For obscure reasons this reason has increased from 6.5% (second half 2000) to 7.4% (first half 2001). The actual reduction is therefore only 9% (6.5% - 5.9%).

Discussion

Until 1998, the drop-out rate at the training battalions was 15% or less. In 1999, the percentage doubled before stabilising in 2000. That represented a significant increase over a short space of time. The search for any important change in that period within the organisation leads to only one that could qualify, the salary cut of \textdollar{}300 (136 Euro) per month that had been implemented several months before. Besides that, or rather, in addition, the economy was very strong during the same period with a very tight labour market. An important characteristic of a tight labour market is that static job satisfaction leads to higher drop-out rates. As a result of the salary cut, there had also been a decline in job satisfaction, however. The coincidence of job dissatisfaction and a tight labour market reinforced both. One must not forget that for the group under discussion here, salary is the single most important aspect of the job.

As a result of the increase of the drop-out rate, various measures were put in place. On the one hand these measures concerned making expectations more realistic by providing realistic information. If (in the eyes of its employees) an organisation has bad terms of employment, however, it will still result in a high drop-out rate and fewer applicants. Therefore, besides ensuring that expectations and reality are better aligned, the organisation has also endeavoured to remedy identified problem areas, with regard to terms of employment, work atmosphere and job content. This took the form of pay rises, the launch of an introduction period and the extension of the General Military Training. The latter two measures were aimed at acquainting trainees better with the RNLA culture and reducing both the mental and physical workloads and creating more time in the evenings to improve the work atmosphere.

The measures resulted in a reduction in all the problem areas identified and an overall decline by 32% in the drop-out rate (from 34% to 23%). This has not yet brought the drop-out rate back down to its original level, but then a number of measures have not had their full effect yet. Furthermore the numbers of trainees who applied were significantly
higher than had been expected, whilst a large number of instructor’s positions were vacant. To remedy the latter, assistant instructors were brought in, but these relatively inexperienced instructors also had to deal with larger class sizes which meant that the courses that required intensive supervision were given less attention than was required. Another problem was that a new way of issuing personal equipment did not run smoothly. The equipment did not arrive on time nor were there sufficient quantities. This led to time loss and frustration for both trainees and instructors. A similar problem occurred with regard to teaching equipment. Although the prognosis is that these problems will resolve they have not been fully taken away yet.

Despite the fact that the drop-out rate has not yet reached its original level, the large intake has resulted in the vacant positions within the RNLA steadily being filled. In addition, the less tightened job market and the improved position of the RNLA on that market compared to that of civilian organisations have meant that more people are willing to extend their contract, keeping positions filled for longer periods of time. The prognosis is therefore that, due to the positive developments with regard to recruitment and retention, the RNLA will in time be fully staffed again.
Figure 2. *Drop-out rate over time*

![Graph showing drop-out rate over time]

*Measures implementation timeline*

Q1 1998 and previous:
- Average Drop-out rate @ 15%

April 1998:
- Salary decrease personnel with fixed contract by EUR 140

2000a:
- Trades assignment in consultation with personnel with fixed contract

2000b:
- Salary increase 17 and 18-year-olds
- Start of Introduction Period
- Preventative transfer to Remedial Platoon

2001a
- Introduction of flexible employment conditions for personnel with fixed contracts
- Release of realistic information video

2001b
- Increase General Military Training from three to four months
- Salary increase
THE LEGACIES OF AN INDUSTRIAL AGE MILITARY CAREER SYSTEM: EXCESSIVE JOB ROTATION AND CAREER TOURNAMENTS

Lieutenant Colonel David Schmidtchen, PhD
Defence Personnel Executive
Australian Defence Force

Introduction

Like all contemporary organisations, the Australian Defence Force (ADF) is a culture in evolution. In its short history it has passed through two main recognisable forms and is now in the throes of a transformation into a third. But, like most military institutions, the ADF is making the transition with some reluctance. It clings to its professional self-image of ‘artisans of war’ and the belief that its considerable professional reputation was derived not despite but because of their small-scale concept of the battlefield.

In the modern era, the maintenance of full-time military forces poses significant resource management challenges to governments. Dealing with the ‘new leadership challenges’ identified in the recent Defence White Paper will require the Australian military to broaden its professional perspective and its concept of its professional role. Increasingly, warfighting will be just one of a number of competencies that make for success in the politically and sociologically complex defence arena.

Any new culture must preserve the best features of the traditional institution. The nature of the military profession might have changed and its essential function of warfighting might have become more complex, but success on the battlefield (however defined in an era of peacekeeping operations) still relies on the skills and values of the warrior. At the same time, however, the ADF must not cling too tightly to tradition in formulating a professional development paradigm to fit the new culture.

All this must be underpinned by a better approach to mid-career selection, streaming and career development. The practice of continual job rotation across a range of staff functions, regardless of an individual’s background and qualifications, belongs to a bygone era when issues changed but slowly and commonsense and general knowledge was usually enough on which to get by. This paper argues, however, the continuance of the ‘enthusiastic amateur’ approach embodied in the organisational practice of excessive job rotation and damaging career tournaments makes no sense in the contemporary era.

These issues are about how the military profession will respond to the challenges of the global era. This is a problem that the profession and its members must confront, not one to be laid at the feet of the ‘institution’ or the ‘system’ in the hope that a solution will emerge.

The Benefits and Costs of Job Rotation

The military continues to be one of the few professions that engages in widespread generalist job rotation as a personal and professional development strategy. In other
professions, people may change jobs just as often but there is invariably a strong degree of continuity or consistency in what they do in successive jobs. Nor do those in other professions continue the practice of job rotation into the middle and senior levels of the professional career. Civilian organisations insist that personnel policy development should be realistically attuned to the need for organisational adaptability and effectiveness. Most thus find it impractical to maintain widespread job rotation when the pace of organisational change is so great.

Figure 1 shows a snapshot of ADF job mobility for selected employment groups up to middle-level officer (major to lieutenant colonel and equivalents). For both officers and soldiers, job tenure increases as the individual advances in rank. However, even for middle-level officers in staff appointments, those in the first year of an appointment outnumber those in the second, and only 16 percent make it through to a third year. Figures 2 and 3 show the current situation for the Defence senior leadership group (ADF and Australian Public Service officers at the rank of brigadier and above). For the military, the current median time in job is one year or less. Their civilian counterparts are not much better. It is highly likely that this degree of mobility adversely affects job performance.

*Figure 1. Appointment tenure in 2000 for selected employment ADF groups.*
Figure 2. Current Senior Leadership Group (SLG) appointment tenure: year of appointment to current position (eg., 9% of military SLG members were appointed to their current position prior to 2000, and 41% were appointed to their current position in 2002).

![Bar chart showing current SLG appointment tenure for each Service: year of appointment to current position.](image)

Figure 3. Current SLG appointment tenure for each Service: year of appointment to current position (eg., 11% of the Navy SLG were appointed to their current position prior to 2000).
Despite this increasing divergence from civilian practices, however, this way of thinking about military careers is so ingrained that policy makers and scholars alike, implicitly or explicitly, take its benefits for granted. Any possible adverse effects are rarely considered. When such drawbacks are acknowledged, this is usually in terms of family issues; but the unspoken assumption tends to be that these are outweighed by career development benefits.

Job rotation does have many benefits but its costs extend beyond just family development effects. Arguably, the Services’ policy of job rotation is such a major impediment to organisational improvement that it is the main sticking-point for other organisational improvement programs. Table 1 sums up the benefits and costs of job rotation as a career development strategy, each of which is then discussed below.

Table 1. Benefits and costs of job rotation as a career development strategy.

<table>
<thead>
<tr>
<th>Year of appointment</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
</tr>
</thead>
<tbody>
<tr>
<td>Navy</td>
<td>11</td>
<td>18</td>
<td>50</td>
</tr>
<tr>
<td>Army</td>
<td>6</td>
<td>31</td>
<td>48</td>
</tr>
<tr>
<td>Air Force</td>
<td>14</td>
<td>21</td>
<td>22</td>
</tr>
</tbody>
</table>

Table 1
<table>
<thead>
<tr>
<th><strong>Benefits of Job Rotation as a Career Development Strategy</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Develops the primary professional self-identity of warrior</td>
</tr>
<tr>
<td>Grooms officers for complex roles at the top of the institution</td>
</tr>
<tr>
<td>Develops the organisation’s capacity to expand rapidly to meet operational liabilities</td>
</tr>
<tr>
<td>Develops an appreciation of the activities and characteristics of related units</td>
</tr>
<tr>
<td>Ensures that both staff work and training will be conducted in empathy</td>
</tr>
<tr>
<td>Contributes to career satisfaction by sharing the ‘good’ and ‘bad’ assignments</td>
</tr>
<tr>
<td>Disruptive to family development</td>
</tr>
<tr>
<td>Costly in terms of geographic relocation</td>
</tr>
<tr>
<td>Over-develops skill in tactical problem solving at the expense of skill in strategic problem solving</td>
</tr>
<tr>
<td>Inhibits job performance, especially in non-primary roles</td>
</tr>
<tr>
<td>Inhibits organisational change</td>
</tr>
<tr>
<td>Inhibits career satisfaction, especially in non-primary roles</td>
</tr>
<tr>
<td>Inhibits the development of a diversity of views at the top of the organisation</td>
</tr>
</tbody>
</table>

The generalist pattern of career development benefits the ADF in many ways. Perhaps the most fundamental of these is that job rotation develops the primary professional self-identity of warrior. Because the only consistency in their careers is that of the primary military speciality, members at all levels come to see themselves as ‘Warriors’ with professional values aligned with the institutional culture.

Secondly, job rotation enables the Services to groom senior officers for what are undeniably complex roles. A modern military institution has hundreds of employment streams and jobs, and is thus equal in complexity to the largest business corporation. Moreover, the modern Defence Force has a number of tasks that it must be capable of performing, usually with little control over what, when, how, where and with whom it will do so. The expansion of the number and complexity of institutional tasks has created a ‘mission revolution’. The executive teams that lead military organisations plainly need better-than-average awareness levels of the component functions and how they interconnect. Because there is virtually no lateral recruitment into the most senior levels, the argument goes, most officers must begin to gain generalist skills in their careers from an early stage, in order to establish the pool from which the future commanders and policy leaders are drawn. Hence all officers tend to be managed as if they have the potential for higher rank, until evidence shows otherwise.

The third benefit of job rotation is that it develops the organisation’s capacity to expand rapidly to meet operational liabilities. The Australian Services have always been characterised by ‘hollowness’, in the sense that their actual organisation is based on a structure that is not a reflection of its true size. In essence, the ADF is a ‘virtual’ organisation, capable of expanding rapidly in times of need. Hollowness, especially for the Army, thus pertains to the maintenance of a latent capability. This capacity to expand rapidly requires officers to be competent in skills that are not necessarily performed in their normal duties. For example, an officer might currently be a captain staff officer but his training will have equipped him to step up to command of a subunit at short notice,
should this be required by mobilisation. One way of giving officers an expanded skill portfolio is to rotate them across a range of different jobs in which they will, at the least, develop an awareness of what they might have to do if such expansion is necessary.

Fourthly, job rotation gives officers an appreciation of the activities and characteristics of related units, and thus an enhanced capability to orchestrate the activities of ‘battle groups’ comprising various elements from different functions. Rotating officers from field positions to the staff and to instructional appointments also allows the Army to ensure that both staff work and training will be conducted in empathy.

Job rotation is also said to contribute to career satisfaction. The Services implicitly recognise the need to share the ‘good’ and ‘bad’ assignments, and there is a desire to give members and their families the variety of experience that has been a traditional benefit of service life. And, finally, the need to fill vacancies is in any case fundamental to an institution that attempts to maintain an ‘optimum’ flow-through of personnel and where prescribed retirement ages are comparatively low.

Costs of Job Rotation as a Career Development Strategy

Despite these many benefits, there are a number of reasons to question the continuation of job rotation, at least in its current form. The practice of high job rotation throughout the career contributes to the stress on service families, leads to expensive service removals budgets, hampers the institution’s capacity for organisational change, limits the contribution a member can make in each job in terms of day-to-day job performance, and creates job dissatisfaction among middle-level officers whose careers have plateaued.

The effect of job rotation on families has been a major issue in service lifestyles for a generation or more. Increasingly, the partners of Service officers are engaged in some form of paid employment, and that their possession of professional qualifications makes it likely that many would be pursuing careers, not just ‘jobs’. However, the rate of geographic mobility of service families has scarcely changed over the last 15 years, with about two in every three having been required to move at least twice in the previous 5 years. In the last 15 years, the number of officers’ partners indicating that they have a good to very good chance of getting an equivalent job if they were to relocate has fallen by about a third (from 65 percent to 41 percent). Similar issues apply to the quality of the education received by children in service families.

The second reason to challenge the continued high rate of job rotation is cost, primarily the cost of geographic relocation. A recent study in the Australian Defence Force found that, in an active duty force of 52,000, 40 percent relocate annually, at an average cost of about $8,000 per relocation. When the costs of providing government housing are included, the average cost per removal more than doubles. Given the increasing pressure on the defence budget, it was inevitable that such costs would eventually be called to account.

The third reason to challenge the continued high rate of job rotation is its effect on job performance. Over 15 years ago, the Careers in Conflict study pointed out that such effects are not trivial, especially for certain professional roles. Careers in Conflict showed that effectiveness in an appointment increases as a function of time in job, with much of the first few months of any appointment being a ‘learning’ phase, after which
effectiveness rises progressively. The study found that performance levels in Service Headquarters and Defence Headquarters staff appointments were consistently below those of their counterparts in operational units. The study suggested, in fact, that many headquarters staff officers really gained little in terms of useful experience during their time in such roles and contributed well below their potential.

A later study confirmed these findings. This study showed that the performance improvement from the first to the second year is often comparatively slight, but those in their third year of an appointment performed substantially better than those in the first and second years. The only exception to this is for middle-level commanders, whose performance levels were high throughout their appointment tenure.

Employment stability generally leads to improved performance among Other Ranks and staff officers. The fewer the postings a person has had in the recent past, the better the current performance. An individual who has had ‘high stability’ (defined as having had only two jobs in the past six years) will, all other things being equal, be a better performer than one who has had low career stability (five or six jobs in the past six years). For regimental officers and middle-level commanders, on the other hand, employment stability was not related to current performance. This suggests that, for officers in these streams, higher rates of job rotation may be justified as an investment in future performance.

In addition, consistently high rates of job rotation negate much of the investment that the Army makes in formal education and other forms of career development. In the first year, qualifications have no effect on performance differentials (that is, the ‘more qualified’ perform at the same level as the ‘less qualified’). In the second year, however, the more qualified are clearly superior; and in the third, the performance gap between the two groups is even wider. Given, however, that only one officer in eight makes it through to this ‘breakthrough’ third year, the Defence Force is plainly foregoing many of the benefits that might otherwise be realised by its investment in intellectual capital.

**Staged Transitions, Tournaments and Diversity**

Those who demonstrate decisive, direct and action-oriented leadership tend to receive favourable performance reports and ratings and tend, therefore, to be promoted to successively higher ranks. The Army’s generalist approach to career development means that promotion criteria reflect the key mainstream military roles at each career stage. Whether lieutenants or brigadiers, officers are judged primarily on their potential for command.

This may be an acceptable approach to career advancement for the first 15 of the officer career, but it has serious disadvantages beyond that career stage. This is because the early career development system, focusing so heavily on the development of tactical problem-solving skills, inhibits the development of strategic skills.

The tactical mode of warfare and training is characterised by problems that require immediate action and centralised control. This often requires the application of standard procedures and ‘drills’, with expectations for prompt, discernible, measurable results. The linkage between cause and effect is clear, and specific performance indicators are usually available for decision making. For such problems, reflection or contemplation is
often inappropriate or unnecessary. Typical mainstream officers fit well into this situation, because they are selected and trained for exactly such a role. After many years practicing this career role, many find it difficult to move from reliance on doctrine and trained skills and procedures to the ‘outside-the-box’ mode of thinking needed for more senior levels.

In the strategic mode, on the other hand, a major feature is the variety and novelty of the issues with which leaders and staff have to deal. Under such circumstances, people are required to think ‘outside-the-box’ as a matter of routine. The role focus at the strategic level is on sustaining or improving the strategic situation, protecting institutional values, reconfiguring organisational systems, investing in basic research and education, and taking time to coach and mentor. The links between cause and effect here are tenuous and outcomes are difficult to quantify, and there are often incomplete or conflicting data from multiple sources. This type of situation gives the mainstream officer ‘more trouble’, according to soldier-scholar General Walt Ulmer, because it requires contemplation before action, patience with ambiguity, consideration of unconventional methods and solutions and an appreciation for broad participation in the decision making process. The more senior the career role, the more the early career style – what Ulmer calls the ‘aggressive adventurer’ – should give way to what he calls the ‘reflective builder’.

This issue of the ‘tactical-strategic transition’ is especially pertinent to the ADF, given its cultural history and the importance of unit-level operations in the professional psyche. The career goal of many officers – perhaps of the majority – is to command at unit level. Lacking a clear concept of what the career might be like beyond that goal, they are strongly inclined to concentrate on skill development for the tactical level of war. And, as Ulmer argues is the case for the US Army, many are victims of ‘over-successful’ development in their early career years, with tactical skills and habits so well-ingrained that they hinder adjustment to the more nebulous later career role. Thus adapting to a different style in mid career is usually a significant challenge on which many officers stumble.

The cultural and performance effects of the strong weight placed on decisive, direct and action-oriented leadership (the ‘heroic leader template’) is complicated by the ‘tournament’ process of career advancement that operates from mid-career onwards. For the first decade or so of their commissioned service, officers advance by cohort, with almost all members of a cohort promoted to captain and again to major at given career points. In the first decade of the career, promotion chance to the next rank is virtually 100% but declines progressively thereafter.

The system in the second half of the career functions as a tournament. This tournament has two main features. First, it is a system for deciding ‘winners’ and ‘losers’. ‘Winners’ are promoted and thereby have priority access to the most important and rewarding career appointments and to career development opportunities in the form of courses and experience (such as higher level staff college and senior field command). The ‘losers’ in successive rounds of the tournament process no longer compete, at least not with those who remain as ‘winners’. Officers who are on the ‘unlikely-to-be-promoted’ list are not only effectively plateaued, but are thereby on the margin of the central core of the ‘warrior’ stream. This gives them much less opportunity to develop core professional skills, especially by experience and practice.
The second feature of a ‘tournament’ is that ‘winners’ progress to the next round, where the process is repeated. Their status as a ‘winner’ lasts only for their current career stage. At the end of that stage, they must compete again, with the bar becoming higher each time.

For a number of reasons, the ADF’s attention focuses mainly on those who remain in the tournament. Especially from brigadier onwards, they are highly visible and on their performance depends, to a large extent, the fortunes both of subordinates and the institution itself. They are the ‘stars’ of the team (at least until the next round of the tournament). Paradoxically, however, the ADF needs to pay equal attention to the so-called ‘losers’, because it is on their foundation that the future performance of the ‘winners’ will depend. Few of those who are passed over from mid level onwards lack competence. Indeed, many will have as much potential for higher rank as those who remain in the tournament. Their handicap is often that the ADF has not come to grips with the way that the career criteria that applied in early career lose much of their validity as the career role shifts from the tactical to the strategic.

Universal high job rotation favours the actioned-oriented type over the ‘reflective builder’ type in the tournament process, because of the limited time available to make one’s mark. Thus officers who fit ‘heroic leader template’ arrive at the top in disproportionate numbers. But the higher a senior commander goes, the more he/she needs the support of the ‘reflective builder’ staff officer/policy analyst whose talents are those of thinking, analysing and communicating. Tragically, there is little attempt to develop the careers and maintain the career commitment of those whose careers plateau, many of whom were passed over simply because they were seen as too ‘reflective’ when they should have been ‘active’. Most of those who drop out of the central core will leave the Army. This is, in fact, what the system wants them to do, to free up promotion opportunities for up-coming officers. In the process, however, the institution unwittingly loses talent that it could usefully employ for much longer and that, increasingly, will be crucial in organisational effectiveness.

Another disadvantage of the tournament process is the way that those who are not ‘insiders’ from the mainstream employments, or who miss one or more tournament events, are critically disadvantaged in terms of competitiveness for higher rank. In the ADF, ‘outsiders’ – even those who, like reservists and specialists, superficially resemble the insiders – are limited in what they are allowed to do and how far they can go. With lateral recruitment being uncommon, because the Services’ unique ways of doing business requires them to ‘grow their own’, the Services’ senior officer/executive teams are probably the least diverse of their equivalents in any Australian organisation. Most are either military officers, who have progressed through the Services by markedly similar experience with markedly similar values, or career public servants. In contrast, it is increasingly common amongst business corporations for the appointment of outsiders to senior jobs, often quite different to the fields from which they came. A recent study of executive career patterns in the US concluded that, since leadership at the head of corporations would become increasingly important and demanding and difficult, those placing executives are likely to favour candidates ‘with diverse backgrounds, rather than more parochial histories’.
Similarly, those who choose to give priority to family development during their early career find it hard – one might say, impossible – to catch up thereafter. This applies especially to female officers, but it might increasingly apply to men as they give temporary priority to their partners’ professional careers. The current system was devised in an age when the ADF’s assumptions about the primacy of the career of the ‘head-of-the-family’ were consistent with community norms. A more flexible approach is needed for an era where lifestyles and family priorities are much more diverse and fluid. In particular, the inflexibility of a cohort-based, ‘staged’ advancement process, which decreasingly suits the more flexible lifestyles of the up-coming generations, is a significant barrier to the creation of greater diversity in the senior teams at the top of the ADF. It means that only those who are prepared to play by the career rules are eligible for each successive tournament. And, whilst this probably is ‘fair’ on an individual basis, it has the serious disadvantage of limiting the opportunities of both women and men who wish to take less conventional approaches to their career. Those in the latter category will be increasingly representative of the Australian mainstream populations, so the maintenance of this kind of career development system is an impediment to both recruiting and retaining those who are capable of presenting the counter view.

The advantages and disadvantages of the tournament approach to career advancement are summarised in Table 2.

Table 2. Advantages and disadvantages of the tournament approach to career advancement.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>?? People advance on merit: on the basis of what they do, rather than who they are</strong></td>
<td><strong>?? Those who choose to give temporary priority to family development during their early career are handicapped in terms of access to career development opportunities</strong></td>
</tr>
<tr>
<td><strong>?? Reinforces the cultural facet of professionalism: the way forward is by developing one’s skills</strong></td>
<td><strong>?? Reinforces a narrow version of professionalism: the way forward is by developing one’s skills in the primary speciality, with other types of expertise counting for little</strong></td>
</tr>
<tr>
<td><strong>?? Reinforces community: in the ‘apprentice’ stages of the career, everyone gets the same opportunities</strong></td>
<td><strong>?? In the later stages of the career, where expertise other than in the core professional function is needed, those with such expertise often lack the organisational authority to get things done efficiently</strong></td>
</tr>
<tr>
<td><strong>?? Reinforces hierarchy: by definition, senior members know more and are usually more competent</strong></td>
<td><strong>?? Reinforces conservatism: almost all members have a stake in keeping the system going</strong></td>
</tr>
<tr>
<td><strong>?? Reinforces conservatism: almost all members have a stake in keeping the system going</strong></td>
<td></td>
</tr>
</tbody>
</table>
Conclusion

The professional military career system is the major conduit by which ADF cultural values are translated into professional behaviour.

The ADF’s current approach to career development is both its strength and its weakness. Its very success in preparing officers for the most visible career role – the Warrior – tends to conceal its limitations. The career system may produce excellent regimental officers and field commanders, but its perpetuation will see continued problems with performance in staff and senior leadership roles, lack of diversity in the senior leadership group, and problems with family development and partner support.

Subtly but profoundly, and in ways both desirable and undesirable, the career system reinforces the ADF’s culture. ADF leaders don’t have to stress the professional standards of their institution, because those standards are implicit in almost every aspect of career development programs. When a system seems to be performing as well as the current ADF career system, there is a natural reluctance to question it too deeply, especially in a conservative culture. Fine-tuning is seen as legitimate: fundamental re-examination is much less acceptable. This, however, is not a wise strategy in a politically and sociologically complex era where organisational paradigms are shifting so quickly and when the combat-oriented image of the military corresponds less and less to what its officers actually do. The ADF’s persistent reluctance to confront this matter is a clear demonstration of how far it has to travel in cultural terms to transition from the current industrial age career development paradigm to one better suited to the global era.

The United Nations, for example, has defined gender mainstreaming as: “the process of assessing the implications for men and for women of any planned action, including legislation, policies or programmes, in all areas and at all levels. It is a strategy for making women’s as well as men’s concerns and experiences an integral dimension of the design, implementation, monitoring and evaluation of policies and programmes in all political, economic and societal spheres so that women and men benefit equally and inequality is not perpetuated” (United Nations, 2000: 1).

The United Nations interest in gender mainstreaming has evolved as a result of changes in the peacekeeping environment. In the past peacekeeping was primarily a military function and included interaction with other military personnel and senior political officials – for the most part, an all-male environment. However, multidimensional peacekeeping operations have evolved to involve many civilians and many women. In relation to peacekeeping, the United Nations has suggested that gender mainstreaming means integrating a gender perspective in the development of mandates, planning and budgeting, leadership, recruitment, training, procedures, and monitoring and accountability (United Nations, 2000). On 31 October 2000, the UN Security Council also adopted UN Resolution 1325 (2000), on ‘Women and peace and security’. This resolution expresses, among other things “its willingness to incorporate a gender perspective into peacekeeping operations, and urges the Secretary-General to ensure that, where appropriate, field operations include a gender component” (UN Resolution 1325 (2000)). “While not using the term ‘gender mainstreaming’, the Canadian government
has had a long-standing commitment to the analysis of policy impacts early in the policy
and decision-making process” (Status of Women Canada).

A final, but important, note is that gender mainstreaming is not intended as a
platform upon which to advocate for ‘women’s issues’, but rather a call for the
integration of the perspectives of a variety of women and men into the development and
conduct of all processes and activities that will impact a diversity of both women and
men. Gender-based analysis is discussed in further detail below as a tool for
implementing ‘gender mainstreaming’.

**GENDER-BASED ANALYSIS**

Gender-Based Analysis is a tool to facilitate the integration of gender perspectives
into the development and design of policy, programs and service delivery. In Canada, for
example, gender-analysis is actively applied to varying degrees in several government
departments including the Canadian International Development Agency (CIDA), Human
Resources Development Canada (HRDC), the Department of Justice, and the Department
of Indian Affairs and Northern Development (DIAND). At CIDA, where gender analysis
is a required aspect of all policies, programs and projects, gender analysis,
“refers to the variety of methods used to understand the relationships between
men and women, their access to resources, their activities, and the constraints they face
relative to each other. Gender analysis provides information that recognizes that gender,
and its relationship with race, ethnicity, culture, class, age, disability, and/or other status,
is important in understanding the different patterns of involvement, behaviour and
activities that women and men have in economic, social and legal structures (CIDA, Oct
02).

Status of Women Canada (2001) has developed a training package on gender-based
analysis which includes a ten step analysis and planning process. The training has been
designed so that it can be tailored to meet the needs of particular mandates and varying
levels of expertise with analysis, and gender and diversity issues. Each step is listed
below along with one example of a question that might be pursued within each step:

1. Preliminary Assessment of Gender Equality Impacts – In what ways are
   both women’s and men’s experiences considered?
2. Outcomes, Goals, Objectives and Indicators – Are there different
   outcomes for women and men and different groups of women and men?
3. Research – How will the research methods ensure that there is awareness
   and questioning of assumptions inherent to the research?
4. Consultation – When should consultations occur and who should be
   consulted?

---

The training package described below (copyright Status of Women Canada) is one example of various
‘gender analysis’ training packages that have been developed independently or customized from the Status
of Women package for use within various departments.
5. Development of Policy Options – How will each option disadvantage some groups or provide advantage for others? Costs and benefits do not fall uniformly on different groups.

6. Making Recommendations – What are the consequences of adopting, or not adopting, gender inclusive options?

7. Communicating the Initiative – Do the identified target audiences reflect the diversity of women and men?

8. Program/Service Design – What will be the ongoing evaluation methods?

9. Program/Service Delivery – What measures are in place to eliminate barriers to access?

10. Evaluation – What are the unintended results of this program?

This training package is available to all federal departments and agencies in Canada. Processes such as this provide a systematic mechanism to ensure that early in a development process, the right questions are asked and answered to ensure that impacts and outcomes in relation to a diversity of women and men and the relationships among and between them are addressed. Gender-based analysis is inclusive in that it considers differences between and among women and men as a result of the nature of relationships between them, and of their different socio-economic circumstances, and also taking into account compounding issues such as race, class, sexual orientation and/or disability (Department of Justice cited by Status of Women Canada). Gender-based analysis is strategic in nature because it assesses impact and outcome for a diversity of women and men – it looks to the future.

The official future that informs policy and program development, whether consciously or unconsciously looks something like this: “There will be a slow but nevertheless inevitable progression toward gender equality” (McCorduck and Ramsey, 1997: 7). Organizations such as the CF also have an ‘official future’ embedded within the organizational culture. However, futurists such as Pamela McCorduck and Nancy Ramsey tell us that the ‘official future’ will never take place and offer scenarios as a tool for developing multiple plausible futures. Scenario-based strategic planning is described below.

**SCENARIO-BASED STRATEGIC PLANNING**

Scenarios have been used for strategic planning in a variety of contexts for some time, and have been recently integrated into the strategy development process for A *Military HR Strategy for 2020* in the CF. The development and use of scenarios is one potential tool for identifying, a range of driving forces, key factors and plausible futures or ‘scenarios’ which can then serve as possible contexts for the development and application of strategic policy and planning. As described by the ‘Global Business Network Canada’, a scenario is “A story, an image, or a map of the future, an interpretation of the present projected into the future, and an internally consistent story about the path from the present to a future horizon”.
The development of scenarios involves a series of steps that facilitate creative and strategic thinking in identifying a potentially endless range of possible forces and factors. Information in reference to potential forces and factors is gathered via scanning and assessment of the external environment, including the domestic and international environments. In the case of the scenarios used for CF strategic HR development, these factors and forces were grouped into two key themes and physically illustrated in relation to one another to represent a continuum of possibilities along the axes. The four resultant quadrants, represent four scenarios, each of which becomes representative of a plausible future context for the application of policy and strategy. It is important to note that scenarios do not represent predictions of the future, but plausible futures that would ideally cover the full range of possibilities.

The scenarios developed to support CF HR strategy development are characterized by degree of economic development juxtaposed against plausible roles within the domestic and global geo-political context, as illustrated in Figure 1, below.
Scenarios such as these represent a macro level picture of the long-term future and have been developed with a particular focus. In this case, the focus was 20 years into the future to facilitate the development of a strategic human resource framework for the CF. Such a framework is then used to guide the development of HR related plans, processes and activities. In the case of the CF, *A Military HR Strategy for 2020* presents a guiding set of HR Principles\(^{49}\) and Strategic Objectives\(^{50}\), along with direction statements accompanying each principle and objective. These principles and objectives have been developed within the context of a broad range of possibilities for the year 2020. Ideally, the scenarios will represent all of the potential possibilities for the year 2020. The CF has also adopted a Long Term Capability Plan (LCTP) for HR to ensure that resource priority is placed on initiatives which satisfy strategic demands.

The next step for the CF is the identification of change objectives for each strategic objective to move us from where we are now, as determined by an internal assessment process, to where we want to go in accordance with our strategic objectives and direction.

---

\(^{49}\) CF HR Principles, presented and expanded upon in *Military HR Strategy 2020* are: Leadership Responsibility, People First, Military Professionalism, Loyalty and Commitment, Well-Being and Morale, Flexibility and Innovation, Fairness and Equity, Professional Development, and Communication and Feedback (National Defence, 2002).

\(^{50}\) Strategic Objectives, also presented and expanded upon in *Military HR Strategy 2020* are: Leadership, Strategic Plan, Culture, Communication, Consultation, Retention, Recruitment, Professional Development, Transition, Health, Well-Being, and HR Systems.
statements. For example, the direction statement for the strategic objective ‘consultation’, reads as follows:

“To be effective the CF must incorporate a consultative process within its structure and culture. Ongoing consultation with a wide network of stakeholders will be the foundation of a transparent HR framework. Through networked communities of practice and mechanisms for feedback, and using relevant tools and technology, the CF will actively seek input and share information with all stakeholders, internal and external. Consultation will be speedy, timely, and conducted in a transparent way such that people will see the processes at work and better understand how and why decisions were made” (National Defence, 2002).

While this direction statement is intended to represent an ‘ideal’ future state, it is also general enough that a broad range of change objectives with varying impacts and outcomes could be identified. How can we increase our certainty that we are identifying and prioritizing the most appropriate change objectives to meet the people challenges of the future?

Scenarios provide some sense of the context or conditions that might challenge policy and programs in the future, but what does this mean for the practitioner? The primary intent in developing scenarios, is to ensure that the activities which are invested in today will move the organization in a direction that will satisfy as many of the conditions offered by the scenarios as possible. That is, ensure that the organization ‘does not put all of its eggs into one future basket’, only to be confronted by an entirely different future. However, once an initiative is approved, how can a policy or program officer ensure that the work that they do to develop those policies and programs have considered a range of potential impacts and outcomes? A potential model for integrating analytical tools such as gender-based analysis in policy and program development, within a strategic framework is described below.

POLICY AND PROGRAM DEVELOPMENT: TRANSLATING STRATEGY INTO EFFECTIVE ACTION

Equipped with an assessment of the external and internal environments, multiple plausible scenarios, direction statements, strategic objectives and a long-term capability planning process, the task is to develop policies and programs, and identify initiatives that will effectively move HR forward from today to the future. As noted in the discussion above the gap between today and tomorrow is framed by two questions:

1. How can we increase our certainty that we are identifying and prioritizing the most appropriate change objectives to meet the people challenges of the future, and

2. Once an initiative is approved, how can we increase the certainty that policy and program development considers an appropriate range of potential impacts and outcomes?

It is suggested here that the integration of a systematic and analytical process would significantly increase the likelihood that the most appropriate objectives are given
priority and that program and policy development will consider an appropriate range of impacts and outcomes. Gender-based analysis (GBA) provides such a tool, and increases the likelihood that potential impact and outcome on all potential stakeholders is considered. Figure 2, below provides an overview of the strategy planning process that has been developed within the CF for HR development, with gender-based analysis integrated into the process.

**Figure 2: The Strategy Planning Process**

![Figure 2: The Strategy Planning Process](image)

**SUMMARY**

The information above has provided an overview of gender mainstreaming, gender-based analysis, including one example of a training plan for gender-based analysis, and the scenario development process. Ideally, the scenario development process provides a full range of plausible futures that could impact organizations and people in a variety of ways. It is suggested that an analytical tool such as gender-based analysis is an essential component of the strategic process as it provides a systematic tool to close the gap between today’s policies and programs and the challenges of the future. Gender-based analysis compliments scenarios by providing policy and program practitioners with a practical analytical tool that provides a picture of potential impact and outcome of policies and programs on people within the context of the future. Finally, gender-based analysis is considered an ideal tool to fill this gap as it is inclusive and strategic. It considers differences between and among a diversity of women and men, and is designed to look ahead through analysis of impact and outcome of policy, program and initiatives on all potential members of the organization in the future.

**BIBLIOGRAPHY**

---

51 Figure adapted from American Management Association, 2002

Canadian International Development Agency:
http://www.acdi-cida.gc.ca/cida_ind.nsf/ (Oct 02)

Department of Justice Canada:


Status of Women Canada, Gender-Based Analysis Information Kit:


Behavioral-Measurement Factor for Enlisted Advancement

Kiwanis G. Short, M.A.
Personnel Psychologist for Naval Advancement Center
Naval Education and Training Professional Development and Technology Center
Pensacola, FL, US

This paper describes a proposal to add a performance-based behavioral measure to the criteria used in selection of junior-level Sailor for promotion. Measuring critical skill competencies is difficult and varies dramatically across occupations. Despite these difficulties, it is quite feasible to integrate such a measure into the existing formula that rank-orders individual Sailors for advancement. The paper will describe our current ‘Final Multiple Score’ system and will propose the addition of a performance-based element that could be ‘customized’ for different Navy occupations (ratings). The criteria for selecting a viable performance-based metric will be reviewed and discussed.

BACKGROUND

As the Navy begins a “revolution in training” aimed at not only ensuring we have the best-trained Sailor possible, but also we assist the Sailor on his or her career progression. Given this new environment it is fitting that we reexamine the Navy’s Enlisted Advancement System (NEAS). At the crux of the system the goal is to promote the most qualified Sailor to the next higher level of leadership. The Navy has successfully used a combination of performance indicators and knowledge examinations for selection since 1949. However, upon reflection of the addition of a ‘customized’ performance-based element added to an already successful system, would only serve increase the credibility and strength of the current system. This paper will explore this proposed initiative.

CURRENT NAVY ENLISTED ADVANCEMENT SYSTEM

The current Navy Enlisted Advancement System is administered in accordance with policy direction from the Chief of Naval Operations (CNO), Commander Navy Personnel Command (COMNAVPERSCOM), and Chief of Naval Education and Training (CNET). Policy documents include OPNAVINST 5450.194, CNETINST 5400.4, and BUPERSINST 1430.16. The advancement quotas are issued by CNO/COMNAVPERSCOM and executed by NETPDT.

Final Multiple Score (FMS)

The advancement selection process uses a Final Multiple Score to determine the rank of enlisted junior-level Sailors already determined fully qualified for advancement. The FMS is considers three main constructs: performance, experience, and knowledge of the individual Sailor. The performance construct includes the Performance Mark Average (PMA), more commonly known as an average of three performance evaluations. The experience construction is reflected by the time-in-rate (TIR) of a Sailor, awards, and
Passed but Not Advanced (PNA) points. And lastly, the Knowledge portion is derived from their Standard Score on the Navy Advancement Rating Examination. Table 1 illustrates specific percentages used to calculate the FMS.

<table>
<thead>
<tr>
<th>US Navy</th>
<th>Final Multiple Score (FMS)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E-4/5</td>
</tr>
<tr>
<td>EXAM*</td>
<td>34%</td>
</tr>
<tr>
<td>PERFORMANCE</td>
<td>36%</td>
</tr>
<tr>
<td>EXPERIENCE</td>
<td>E4-E6 Only</td>
</tr>
<tr>
<td>TIR</td>
<td>13.0%</td>
</tr>
<tr>
<td>Awards</td>
<td>4.0%</td>
</tr>
<tr>
<td>Passed Not Adv-PNA</td>
<td>13.0%</td>
</tr>
</tbody>
</table>

Table 1

Final Multiple Score for E-4/5/6/

Note. Adapted from data located in the Navy Advancement Manual, BUPERSINST 1430.16.

Example

For the purpose of this paper we will use the E-5 paygrade as an example. Figure 1 below illustrates the breakdown of the E-5 FMS in a pie chart format.

Figure 1. FMS for the E-5 paygrade.
Contribution to Variance

Upon examining the distribution of each component we find that the Knowledge/Rating Examination and the Performance Mark Average/Evaluations are contributing the more variance to the FMS than originally planned by Naval leadership. Figure 2 is an example of the E-5 FMS for Aerographer rating in the Navy for one performance cycle.

![Contribution to Variance](image)

Figure 2. Contribution to variance for the E-5 Aerographer, Naval Advancement Center 2002.

Contribution of these two main scores is a direct reflection of the distributions created by each element in the Final Multiple Score. When the distribution does not follow the normal curve frequently the scores “do not contribute” as designed. Sometimes this is due to system contamination or environmental influence. For example, Awards will contribute less to the FMS if “everyone” is receiving an award. That is, if the award guidelines are not strictly followed and all potential candidates have the same number of awards, the contribution becomes null. However, in the same example there is a correlation between the number of awards a Sailor as received and the number of years he or she has served in the military, which also affects the contribution this factor will have to the FMS.

One of the governing guidelines of the current performance evaluation system is an imposed “forced” distribution of performance evaluation scores. For example, each command officer is limited to the number of high scores he or she can give to one particular paygrade. Since this change was implemented it has significantly increased the
contribution to variance provided by the Sailor’s performance mark average. However, there is a proposal within the Navy community to eliminate the forced distribution guidelines for performance evaluations. If this occurs it will give even more merit to the need for a behavioral-measurement factor proposed in this paper.

PROPOSAL

Our proposal to add a behavioral-measurement to the NEAS will serve to increase the impact and correlation performance will have on the advancement of the Sailor. It will also serve to increase face, content, and construct, and validity of the NEAS. This will be visible not only to the Naval leadership, but to the individual Sailor.

![Pie chart showing distribution of factors]

Figure 3. Proposed change to FMS. Note: Percentages may change as additional research is completed in this arena.

HISTORY OF MEASUREMENT

It is important that we review the past prior to changing the future. Much of the testing for the military began with oral trade tests developed in the Army in World War I. Volunteers reported they had skilled expertise, but it was soon discovered the self-report and requirements were often significantly different. This is when oral trade tests developed. Volunteers would answer questions about particular job skills. Note tests were oral because many still could not read. This evolved into traditional paper pencil testing. This type of testing was standardized, reliable, and could be administered to large groups of people (Guion, 1998).
The basis for any type of behavioral measurement should begin with an occupational analysis. For the Navy occupational analyses are completed by Navy Manpower Analysis Center (NAVMAC) in Millington, TN. A new method of large-scale occupational analysis is being used by the United States Department of Labor (Peterson, Mumford, Borman, Jeanneret, & Fleshman, 1995) and DoL guidelines are followed by NAVMAC. It is a survey method; designed to measure several kinds of job-relevant information. In this work, the model for occupational analysis is referred to as O*NET (for Occupational Information Network). The model is intended for comparing jobs and occupations. O*NET consists of six major categories: Experience Requirements, Occupational Requirements, Occupation-Specific Requirements, Occupation Characteristics, Worker Characteristics, and Worker Requirements. Once an in-depth occupational analysis has been completed a behavioral-measurement factor can be proposed.

METRIC CRITERIA

The most difficult is determining a methodology to measure critical skill competencies, especially since competencies vary dramatically across occupations. In our example, the rating community could assist in identifying critical competencies needed by an Aerographer (the Navy’s meteorologist). However, as you can image the critical skills for a Machinist Mate would be dramatically different.

The criteria listed in Table 2 are recommended to evaluate and certify any type of behavioral-measurement factor prior to being added to the FMS.

Table 2. Criteria for a behavioral-measurement factor

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid-job related factor</td>
<td>Provides face, content, and construct validity</td>
</tr>
<tr>
<td>Correlation to important job outcomes</td>
<td>Provides content validity</td>
</tr>
<tr>
<td>Reliable</td>
<td></td>
</tr>
<tr>
<td>Contain variability</td>
<td>Provides variance among qualified candidates</td>
</tr>
<tr>
<td>Submeasures (7)</td>
<td>Provide construct validity, submeasure limits important to process understanding</td>
</tr>
<tr>
<td>Scale of 100 points</td>
<td>Easily understandable and variable</td>
</tr>
<tr>
<td>Descriptive and understandable</td>
<td>Provides everyone an equal opportunity and visibility of the selection system</td>
</tr>
<tr>
<td>Objective measurement preferred</td>
<td>Limits subjectivity, conveys fairness</td>
</tr>
<tr>
<td>Subjective measurement guidelines</td>
<td>Provides for inter rater reliability</td>
</tr>
<tr>
<td>Fleet/Shore Equity</td>
<td>Conveys fairness to every candidate</td>
</tr>
<tr>
<td>Compromise Risk should be minimum</td>
<td>Essential to proper selection and credibility of system</td>
</tr>
<tr>
<td>EEOC validation</td>
<td>Legal stability, meets Title VII standards, and shows job-relatedness</td>
</tr>
<tr>
<td>No disparate impact</td>
<td>Legal implications and fairness to all Sailors</td>
</tr>
</tbody>
</table>
Note: Table is created based on primarily expert knowledge and information provided in a variety of sources. Brown (2000) provides a useful guide of questions to be used when designing metrics. Flynn and Gillian (2002) explore the legal issues of testing and the need to complete an EEOC validation when impacting the selection of personnel.

After the criteria has been chosen the requirements for a valid measurement need to be addressed. For example, it is necessary to create an established baseline, method of data collection, frequency of measurement, and reporting methodology. Most of these requirements reflect good research and implementation techniques, however, they are also necessary to provide a solid legal foundation for the system.

**DISCUSSION**

In summary if the US Navy selects, tests, and implements a behavioral-measurement component to the FMS for junior-level Sailors it would be very beneficial. The very nature of a behavioral component would add face, content, and construct validity to a sound selection system. However, the selection and implementation of a behavioral-measurement is a very tedious process. Because there are over 85 occupational ratings in the Navy the measurement selection process could be very time consuming. Fortunately, the concept could be implemented in stages since one rating’s behavioral-measurement would not affect the selection of personnel in other ratings.
REFERENCES


Why Smart Classification Does Matter

LtCol Psych Francois J. LESCREVE
Belgian Defense Staff
Human Resources Directorate General
Accession Policy Research Section

1. Introduction

In the process of recruiting, different phases can be distinguished from a conceptual point of view. In a multiple-applicant, multiple-job environment, two distinct phases are always present. One phase is focused on assessing the applicant. Through tests, interviews, physical examination, questionnaires and other selection tools, pertinent information about the applicant is gathered. The other phase is about decision-making. Does the applicant get a job or not? What job does he or she get? This decision making process is referred to as the ‘classification’ of the candidates. For a given set of applicants and jobs, numerous solutions to allocate the persons to the jobs are conceivable. The paper will demonstrate that the choice of the used classification methodology has a significant influence upon the level of job specific aptitudes that can be expected within the groups of persons that are assigned to the different entries.

Among the possible classification strategies, two broad families can be identified. The first family aims at immediate decision-making. That usually means that a decision about the allocation of a person to a vacancy is done while the candidate still is at the selection center, disregarding the aptitudes and other relevant attributes of the following applicants. The second family on the other hand is referred to as ‘batch classification’. This approach compares all persons in the applicant pool before making decisions. It will be shown that the second type of approach produces far better results.

This paper reviews a number of alternative classification strategies and illustrates by means of recent empirical research results how the chosen strategy affects the obtained average aptitudes in the different entries. Quantitative arguments are given why the personnel manager concerned with recruit quality, should choose more complex classification methodologies referred to as ‘smart classification’.

2. Classification

Imagine that one lucky morning, your boss asks you in his office and tells you: “I’m quite happy with the work you do, but I noticed that you spend too much time doing purely administrative tasks. I therefore decided to give you a secretary so that you can concentrate on more important parts of your job. I contacted an employment agency and they have three candidates. You’re free to choose who you want.” How will you proceed? If we formalize this situation a bit, we can say that there is a job and three applicants for that job. Obviously you would like to choose the person that is best fit for the job. To find out who that is, you’ll probably start thinking about what aptitudes, skills, interests and so on are relevant for this new job and how important each of these attributes are to perform well. More formally, you
would conduct a job analysis to identify the relevant attributes and by giving these attributes a relative importance or weight, you would define a job profile. As a next action, you would see the three applicants for the new position and try to find out to what extent they possess the requested attributes. In other words, you would assess the relevant attributes of the applicants and by doing so, define their profile. Finally, you would have to decide which applicant to hire. If you have the applicants’ profiles and know the relative importance of each attribute for the job of secretary, you can appraise how appropriate it would be to hire each one of the applicants. This can be referred to as computing the utility of assigning each applicant to the job. You then obviously would choose the person showing the largest utility.

When we transpose our simple example to the highly complex multiple-applicant, multiple-job situations that we encounter in Military recruiting settings for example, the logic remains essentially the same. For each job, the relevant attributes must be identified and their relative importance is to be known. On the applicants’ side, each of them needs to go through an assessment system that will evaluate his or her aptitudes and interests and define an individual profile. Based upon the individuals’ profiles and the jobs profiles, the utility of assigning each person to each vacant job can be computed. Now however, a major difference distinguishes the multiple-job situations from the simple example we gave earlier. In our example, the decision was straightforward as soon as we had computed the utilities of hiring each one of the three applicants. The person with the highest utility would get the job and that’s it. In the more complex multiple-job situation, the decision-making is more challenging. Typically, the required attributes and their relative importance will differ for each trade. And whereas an individual’s skills and aptitudes are essentially independent of the examined job, his or her degree of interest for each kind of job usually varies. This means that the utilities we’re interested in may vary from person to person and for each person, from job to job. When we consider a group of vacancies and the applicant pool applying for at least one of the vacant jobs, we could represent the decision-making problem by means of a matrix. We could have the jobs as column headers and the persons as row headers. Each cell would then contain the utility of assigning the row-person to the column-job. The task at hand is then to link persons and jobs. That is referred to as classification.
The number of possibilities to assign applicants to jobs rapidly becomes astronomical as the number of jobs and applicants increase and that makes the problem both challenging and interesting. The interesting part comes from the fact that this problem has lots of degrees of freedom and that it makes sense to investigate the relationship between the method used to solve the classification problem and the quality of the reached solution. The remainder of this paper will discuss a few methods that are used to solve the classification problem and illustrate how the chosen method affects the hired group. The examined methods are:

- Immediate classification;
- Single criterion batch classification;
- Batch classification based upon multiple rank order criteria (parallel processing);
- Smart classification, i.e. batch classification based upon multiple rank order criteria (optimization algorithm).

The first method, called ‘immediate classification’, assigns an applicant as soon as all his relevant attributes are assessed. In order to decide about his assignment, his ‘profile’ is compared to a set of trade specific criteria. If the person meets the set criteria for his preferred job, he gets it. This system, also known as a ‘first comes, first served’ system is widespread for enlisted personnel. The main reason for this is that it usually is considered to be important to tell the applicant immediately what job he will get.

The second method that we will refer to as ‘single criterion classification’ is based upon a single rank order criterion. In this method, all applicants are assessed first. They subsequently are rank ordered by their score on one (composite) score. For instance, the Belgian Royal Military Academy sorts the applicants according to a single composite score made of maths along with first and second national language. Then the applicants are processed sequentially, starting with the best-ranked person. The preferences of the applicant will then be examined sequentially and the applicant will get the job he prefers if he meets the other eligibility criteria and there still is a vacancy available for the chosen trade.

The third method, called ‘multiple criterion classification’ is again a batch classification method. Now however, a ranking is made for each trade separately, based upon trade specific weighted criteria. The classification can for instance be obtained in following way:

For each trade a table is made. The tables contain three fields: the person identification, the payoff or utility assigning the person to the trade and the rank of their preference for the table-entry (a value 1 indicates that the entry is the first choice of the applicant, 2 is his/her second choice, etc). The tables are sorted in descending order of the payoff. The method then examines the number of vacancies per entry. If n is the number of vacancies, then the method will assign persons to the trade if they are among the n best ranked persons for the entry and the entry is their first choice. This would be done for all entries. Once a person is assigned to a job, his or her record is deleted in the other tables. This causes shifts in the tables, potentially leading to new assignments. This procedure is continued as long as persons can be assigned to their first choice. Then, after adapting the number of vacancies

---

2 for the social and military sciences faculty
3 the described method is called the ‘sequential parallel assignment method’ and was proposed by the Belgian Royal Military Academy.
per entry (set to the original number minus the number of persons assigned to the entry in the run examining first choices), the second choice is examined in a similar way. Then the third choice is reviewed and so on. The method stops as soon as there are no more vacancies or all choices have been reviewed.

The fourth and last method is the so-called ‘smart classification’. This is a batch classification method in which the aptitude estimate based upon trade specific weighted criteria and the preference for an entry are integrated into a single utility value for each possible person-job combination. Then, these values are organized in a single matrix featuring the persons as row headers and the jobs as column headers. Next, the persons are linked to jobs using an optimization algorithm that maximizes the sum of utilities for the group of assigned persons.

In order to assess the outcome of the different classification methods, we took an existing dataset describing Belgian NCO’s. The dataset encompasses 393 eligible applicants for 22 trades and a total of 94 vacancies. The measures shown in the next graph are the average scores for different selection variables of the persons that were assigned by the different classification methods.

[Diagram of Mean scores of hired applicants as a function of used classification method]

4 In this method, it can happen that an applicant is assigned to his/her second choice for instance because at that time s/he doesn’t qualify for the entry of his/her first choice. If, during the classification process, due to the deletion of the records of persons who were assigned to an entry of a higher choice, this person’s first choice becomes available for him/her, then the person will be assigned to his/her first choice while the vacancy of his/her second choice will be made available again for other persons.

5 The described method is called ‘The Psychometric Model’ and is currently in use for the Belgian Defense.

6 Recruitment for francophone NCO’s, level 2 in 2000.

7 By ‘eligible’ we mean applicants that meet all legal requirements and therefore can be assigned to a job as NCO.
First the bottom line is given as a benchmark. It represents the (standardized) averages for all persons in the applicant pool. The next line (with open squares) is based upon the immediate classification. The third (dots with open diamonds) is produced by the single criterion classification. In this example, the criterion that was used was the intelligence score. So it isn’t surprising to see the very high average intelligence score produced by this method. It is interesting to note that the averages of correlating aptitudes (mechanics and electricity) benefit from this method while these of uncorrelated attributes (personality and physical fitness) don’t. The fourth line (open triangles) originates from the multiple criterion classification. This method produces a more balanced profile. Finally, the last line (filled circles) comes from the smart classification. The line shows a balanced solution that is markedly better than the one produced by the immediate classification and slightly better that the multiple criterion classification.

In assessing the outcome of the different classification methods, it is also useful to look at some other results. The following graph presents the average utilities, how well the applicants’ preferences are respected and the fill rate.

<table>
<thead>
<tr>
<th>Classification methods</th>
<th>Average Payoff(R)</th>
<th>Average Preference(L)</th>
<th>Number assigned(L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate</td>
<td>82</td>
<td>99</td>
<td>500</td>
</tr>
<tr>
<td>Sequential single crit.</td>
<td>84</td>
<td>98</td>
<td>520</td>
</tr>
<tr>
<td>Sequential multi crit.</td>
<td>86</td>
<td>97</td>
<td>540</td>
</tr>
<tr>
<td>Smart classification</td>
<td>88</td>
<td>96</td>
<td>560</td>
</tr>
</tbody>
</table>

When looking at the average utility (payoff) of the assigned persons, it is quite clear that an increased complexity of the classification method results in an improved quality. This also seems to be the case with the respect of the applicants’ preferences. In our dataset, the applicants expressed their preference for each trade on a scale ranging from 1 to 99. 99 is the highest value and is given to the most preferred trade. The graph presents the average of the preference of the applicants for the trade they were assigned to. An increasing average reflects a higher degree of satisfaction in the hired group. The third line, representing the number of assigned persons, also puts the preference line in a broader perspective. For the

---

8 Although this doesn’t entirely show in this example, unpublished research by the author based upon a large number of datasets demonstrates this statement.
preferences, we see that the average resulting from the multiple criteria classification is somewhat higher than the one produced by the smart classification. This is not in accordance with findings based upon other datasets demonstrating that the smart classification usually produces the highest degree of satisfaction. The reason why this doesn’t occur in this dataset might be due to the fact that the multiple criteria classification only succeeded to fill 88 out of 94 vacancies. In such circumstances, the comparison of mean preferences is biased and overestimates the mean preferences obtained with the multiple criteria classification. As one can see, only the smart classification was able to fill all vacancies. The reason why the sequential methods couldn’t is due to the fact that for some trades, only a few applicants qualify. By classifying them sequentially, the methods assign them without taking into account that this might create a problem for more critical trades. Only the smart classification considers all trades and all applicants globally and has the necessary flexibility to solve the problem.

To complete the assessment of the outcome of classification methods, we propose a second example. This dataset originates from the US Air Force. It contains all enlisted personnel in 1996 and that represents 23578 persons for exactly the same number of positions. The jobs are grouped in 146 different trades. For this example, we compare the original outcome i.e. the jobs the applicants actually were assigned to by the immediate classification with the jobs that the applicants would have got if the smart classification had been used. So we compare the outcome of two classification systems working with exactly the same applicants, the same vacancies and the same criteria defining trade eligibility. The following graphs show each two score distributions. The left distribution represents composite scores (respectively Mechanical, Admin, General and Electronics) obtained by the immediate classification and the right one represents the same scores when the smart classification is used. The persons from who the scores are included are these applicants that have been assigned to a trade for which the considered composite score is relevant.

9 The used method is the Belgian Psychometric Model.
10 This means that in order to qualify for the trade, the applicant has to meet some conditions including the considered composite score. These conditions come under three modes: (1) a composite score larger than a cut off must be reached, (2) a composite score larger than a cut off must be reached AND another composite score also needs to reach a certain level, and (3) a composite score larger than a cut off must be reached OR another composite score needs to reach a certain level.
As can be seen, the smart classification method outperforms the immediate classification clearly. A word of caution is needed however. The used dataset didn’t include the applicants’ preferences nor could we include rules related to minority or female fill rate. It is believed that the inclusion of these elements would have caused a shift to the left of the composite score distribution obtained by the smart classification. On the other hand it is also highly unlikely that these elements could explain the whole magnitude of the observed shift in quality. To put these results into perspective, it would be interesting to consider what alternative actions would be needed to produce a similar improvement of recruit quality. The only solution we see is to foster an improvement of the quality of the applicant pool. In order to reach that under unchanged labor market circumstances, it would most likely be necessary to implement very expensive programs including more advertising, more recruiters and higher recruitment bonuses and wages. In comparison, smart classification looks like having a very appealing return on investment.

3. Discussion and conclusion

To discuss the merits of the different classification methods, it might be useful to elaborate on following table.

<table>
<thead>
<tr>
<th></th>
<th>Fairness</th>
<th>Recruit quality</th>
<th>Respect of preferences</th>
<th>Fill rate</th>
<th>Flexibility</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate classification</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Single criterion classification</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Multiple criteria classification (parallel processing)</td>
<td>+ +</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Smart classification (optimization algorithm)</td>
<td>+</td>
<td>+ +</td>
<td>+</td>
<td>++</td>
<td>++</td>
<td></td>
</tr>
</tbody>
</table>

If we consider the fairness of the methods, we must state that the immediate classification isn’t a fair one. To illustrate this, let’s go back to the example of the recruiting of a secretary. We had three applicants for the job. Well, the immediate classification method would assess
the suitability of the first applicant and make a decision to hire this person or not before the
other two applicants had the opportunity to demonstrate their suitability. Just imagine that
you are the second or third applicant and that you’re told that the position was given to the
first applicant. Hardly fair, isn’t it? The relative unfairness of the single criterion
classification results from the fact that only one (composite) criterion is used to define the
order in which the applicants are reviewed for assignment. When that criterion doesn’t relate
to the trade you’re applying for, it’s not really fair. For instance, if you’re very good at
sports and apply for a job as infantryman, you probably won’t be happy with a classification
system that only will look at your intelligence test score to define the place you get in the
assignment sequence. Also the smart classification can be criticized for its fairness. The
reason for this lays in the fact that the classification focusses on the group level rather than
on the individual level. This can result in the fact that best ranked applicants for a specific
trade that they prefer do not necessarily get that trade if the overall solution can be improved
by assigning the person to a trade of less preference or aptitude. In practice however this
very seldom occurs for exactly the same reason. The classification algorithm looks for well-
ranked applicants who like the trade and so will be attracted to assigning these applicants to
the trade they ask.

The difference in recruit quality obtained by the examined classification methods was
illustrated in our examples. The smart classification produced the best results. It is actually
important to notice that the quality of the hired group, not only is depending on the qualities
available in the applicant pool but also on the used classification method.

The respect of the applicants’ preferences is poor in the immediate and single criterion
classification methods. The reason why is that these methods work in a strict sequential
mode. That usually means that the first persons considered for assignment will get their
preferred job but at the end of the classification process, the story is totally different. There,
when the number and the popularity of available vacancies shrinks, the classification
methods need to assign persons to trades they’re really not interested in and this lowers the
overall satisfaction of the hired group.

When we look at the fill rate, it is clear that only the smart classification method has the
necessary degrees of freedom to guarantee the highest possible fill rate. In times where it is
difficult to reach recruitment goals, this is an important facet of the evaluation of the
classification methods.

In terms of flexibility, the smart classification outperforms the other methods by far. Only
batch classification methods can react quickly to occurring changes in the selection ratio.
And in comparison to the multiple criteria classification method used in our example, we
can say that this method does not allow preferences measured at the interval level, nor the
use of categorical data nor allows the integration of the organization’s priorities. All that can
easily be done in our smart classification model.
In summary, we think that the smart classification method is the best option when classification (or even allocation) decisions need to be made. The immediate classification is widespread for enlisted personnel nowadays and that is because it is thought to be important that applicants get the information about their assignment without delay. We easily can see the advantages of such an approach. But it is equally important to realize that the use of immediate classification has a very high price in terms of recruit quality and overall satisfaction of the hired group. We therefore encourage responsible managers in recruiting organizations to investigate the price they currently pay for using less powerful classification methods. This can rather easily be done by simulating the outcomes of different classification methods when applied to the same dataset.

4. References


• Robertson I., Callinan M. & Bartram D. *Organizational Effectiveness: The role of Psychology*. New York, John Wiley & Sons Ltd, 2002

<table>
<thead>
<tr>
<th>Author</th>
<th>Pages</th>
<th>Author</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggarwal, S.</td>
<td>153</td>
<td>Gutknecht, S.</td>
<td>208</td>
</tr>
<tr>
<td>Alderton, D.L.</td>
<td>71, 76, 411, 554, 557</td>
<td>Halpin, S.M.</td>
<td>263</td>
</tr>
<tr>
<td>Algra, G.E.</td>
<td>493</td>
<td>Hedge, J.W.</td>
<td>416, 562, 588</td>
</tr>
<tr>
<td>Andre, T.</td>
<td>257</td>
<td>Heffner, T.S.</td>
<td>161, 528</td>
</tr>
<tr>
<td>Annen, H.</td>
<td>208</td>
<td>Held, J.D.</td>
<td>291</td>
</tr>
<tr>
<td>Baisden, A.G.</td>
<td>201</td>
<td>Helba, C.V.</td>
<td>331</td>
</tr>
<tr>
<td>Bana e Costa, C.</td>
<td>511</td>
<td>Helmick, J.C.</td>
<td>331</td>
</tr>
<tr>
<td>Bearden, R.M.</td>
<td>416, 562</td>
<td>Hess, K.</td>
<td>541</td>
</tr>
<tr>
<td>Belanich, J.</td>
<td>22</td>
<td>Hetrick, C.</td>
<td>257</td>
</tr>
<tr>
<td>Bender, P.</td>
<td>322</td>
<td>Hindelang, R.</td>
<td>461, 554</td>
</tr>
<tr>
<td>Bennett, N.</td>
<td>120</td>
<td>Jenkins, D.</td>
<td>283</td>
</tr>
<tr>
<td>Blackstone, T.F.</td>
<td>566</td>
<td>Johnston, B.F.</td>
<td>153</td>
</tr>
<tr>
<td>Blanco, T.A.</td>
<td>291</td>
<td>Kammrath, J.L.</td>
<td>184</td>
</tr>
<tr>
<td>Boerstler, R.E.</td>
<td>184</td>
<td>Karrasch, A.</td>
<td>467</td>
</tr>
<tr>
<td>Bourne, M.J.</td>
<td>588</td>
<td>Kaufman, J.D.</td>
<td>562</td>
</tr>
<tr>
<td>Borman , W.C.</td>
<td>416, 562</td>
<td>Kilcullen, K.</td>
<td>176</td>
</tr>
<tr>
<td>Brady, E.J.</td>
<td>188, 235</td>
<td>Klammer, J.</td>
<td>52</td>
</tr>
<tr>
<td>Branson, R.K.</td>
<td>46</td>
<td>Knapp, D.</td>
<td>37, 161, 452, 528</td>
</tr>
<tr>
<td>Brown, K.J.</td>
<td>153, 242</td>
<td>Knudson, K.H.</td>
<td>399</td>
</tr>
<tr>
<td>Brown, P.</td>
<td>221</td>
<td>Kubisiak, U.C.</td>
<td>588</td>
</tr>
<tr>
<td>Bruskiewicz, K.T.</td>
<td>588</td>
<td>Leitão, M</td>
<td>511</td>
</tr>
<tr>
<td>Clermont, Y.</td>
<td>8</td>
<td>Lescreve, F.</td>
<td>643</td>
</tr>
<tr>
<td>Cole, A.K.</td>
<td>153</td>
<td>Levey, M.J.</td>
<td>576</td>
</tr>
<tr>
<td>Collin, I.A.</td>
<td>322</td>
<td>Lickteig, C.W.</td>
<td>191</td>
</tr>
<tr>
<td>Collins, N.</td>
<td>221</td>
<td>Macdonald, G.</td>
<td>1</td>
</tr>
<tr>
<td>Cooper, G.</td>
<td>257</td>
<td>Matser, R.</td>
<td>596</td>
</tr>
<tr>
<td>Cotton, A.J.</td>
<td>570, 576, 581</td>
<td>McCann, C.</td>
<td>263</td>
</tr>
<tr>
<td>Couture, C.</td>
<td>4</td>
<td>McConne, A.</td>
<td>140</td>
</tr>
<tr>
<td>Creamer, M.</td>
<td>576</td>
<td>Meijer, M.</td>
<td>493</td>
</tr>
<tr>
<td>Crookenden, M.P.</td>
<td>291</td>
<td>Michael, P.G.</td>
<td>554</td>
</tr>
<tr>
<td>Curnow, C.K.</td>
<td>22</td>
<td>Miller, P.</td>
<td>541</td>
</tr>
<tr>
<td>Daley, K.E.</td>
<td>82</td>
<td>Modjeski, R.B.</td>
<td>488</td>
</tr>
<tr>
<td>Dewar, M.E.</td>
<td>82</td>
<td>Mottern</td>
<td>76, 557, 560</td>
</tr>
<tr>
<td>Donohue, J.J.</td>
<td>342</td>
<td>Mylly, J.</td>
<td>536</td>
</tr>
<tr>
<td>Entin, E.E.</td>
<td>541</td>
<td>Newell, C.E.</td>
<td>461, 482</td>
</tr>
<tr>
<td>Ervin, K.S.</td>
<td>215</td>
<td>Olmsted, M.G.</td>
<td>475</td>
</tr>
<tr>
<td>Fancher, P.</td>
<td>399</td>
<td>Perry, S.</td>
<td>331</td>
</tr>
<tr>
<td>Farmer, W.L.</td>
<td>411, 416, 554, 560, 562</td>
<td>Pigeau, R.</td>
<td>263</td>
</tr>
<tr>
<td>Fedak, G.E.</td>
<td>291</td>
<td>Putka, D.J.</td>
<td>28, 37, 452</td>
</tr>
<tr>
<td>Ferstl, K.L.</td>
<td>562</td>
<td>Riedel, S.L.</td>
<td>467</td>
</tr>
<tr>
<td>Fleck, G.</td>
<td>250</td>
<td>Rosen, L.N.</td>
<td>399</td>
</tr>
<tr>
<td>Freean, M.W.</td>
<td>22</td>
<td>Rosenfeld, P.</td>
<td>461, 482</td>
</tr>
<tr>
<td>Gogol, O.</td>
<td>446</td>
<td>Rumsey, M.G.</td>
<td>62</td>
</tr>
<tr>
<td>Name</td>
<td>Page(s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rymsza, B.</td>
<td>201</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sager, S.E.</td>
<td>28, 452</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sanders, W.R.</td>
<td>191</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schmidtchen, D.</td>
<td>618</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schwerin, M.J.</td>
<td>588</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short, K.G.</td>
<td>636</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spector, A.</td>
<td>446</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stanley, P.</td>
<td>113</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steinberg, A.G.</td>
<td>442</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stemate, L.</td>
<td>434</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanner, L.M.</td>
<td>52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taylor, I.</td>
<td>434</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Truhon, S.A.</td>
<td>425</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Underhill, C.</td>
<td>475</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vanderpool, M.A.</td>
<td>93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>van de Ven, C.P.H.W.</td>
<td>298, 596</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wait, T.</td>
<td>366, 386</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waldkoetter, R.O.</td>
<td>272, 488</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walsh, K.</td>
<td>257</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waugh, G.W.</td>
<td>28, 37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watson, S.E.</td>
<td>416, 554</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weerts, J.M.P.</td>
<td>493</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wheeler, J.</td>
<td>105</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White, L.A.</td>
<td>169</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White, M.A.</td>
<td>76, 557, 560</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wickes, A.</td>
<td>257</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wilcove, G.L.</td>
<td>420</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Williams, P.M.</td>
<td>566</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>