

## MILITARY PERSONNEL RESEARCH: AN HISTORICAL PERSPECTIVE

Leland D. Brokaw  
 Personnel Research Laboratory  
 Department of the Air Force

Because so many members of the audience have spent major and minor segments of their professional life dealing with the nuts and bolts of the design phases of the present structure, this paper will not dwell upon the specifics of the development of a test, the rationale for a psychomotor apparatus, or the procedures followed in the collection of critical incidents. The evolution of types of research will be traced in the hope of making more clear the enduring pressures and priorities that underlie military personnel research.

It is interesting to attempt to set the date from which this review should proceed. This is the Twenty-fifth Anniversary Symposium; and it is quite true that November 1941 is an historic date in the United States Army Air Corps Aviation Psychology Program. To imply that our present circumstance had its origin at that precise instant would be misleading. It is not appropriate to set the zero point at the time that Dr. Yerkes and his colleagues assisted the War Department with the testing program leading to Army Alpha in World War I, because they depended upon still earlier researchers in the domain of mental measurement. Wilhelm Wundt, and the psycho-physiologists of his day, in a sense mark the beginning; it is improbable that profit would come from the identification of the precise instant and place whence sprung our current program. Better far that we should treat it as a mental measurement scale, identify the origin as undefined, and attempt to work along from the arbitrary point established by this birthday party.

This review will be biased. Since 1943 the writer has worked within the Aviation Psychology Program, in those organizations that have evolved into the present Air Force Personnel Research Laboratory. Formal sources for the material of this review could be the program documentation, the technical reports, and the papers we have placed in the journals, and at the professional meetings. However, this discussion will reflect experience; the program documents, the technical reports, the journal articles, and APA papers are like the exposed portions of icebergs—indications of general shape and size, but concealing major portions of interest.

Investigation of the characteristics of the published materials of the Laboratory reveal some continuity of certain program aspects, certain cycles of investigation that ebb and flow, and isolated explosions of specific activity.

There is a steady tom-tom beat of production, calibration, and validation of operational tests. Indeed the tap root of personnel research springs from the exploitation of individual differences in selection and classification programs.



The dramatic relations between stanine level and elimination rates for pilot trainees who had been screened on an aptitudinal basis in World War II are well documented, well regarded, and well worn as justification for other projects in the personnel research area. Incidentally, the matter of pilot selection continues. We find the same types of paper-and-pencil tests to be valid now that were valid in 1942, and that the same principles of supply and demand apply.

The considerations which led to the use of the aircrew measures provided a basis for the development of relatively sophisticated tests to use in the selection of airmen for service in the Air Force and to define their area of occupational assignment within the Air Force. These measures demonstrated very useful levels of validity against criteria of school success, and here we find the eternal question in personnel research. Our field of significant achievement is also our field of significant need. Our tests have had solid success in the identification of those individuals who might be expected to do well in training, yet attempts to relate test scores to operational measures of success on a job have yielded insignificant results.

This dilemma, recognized throughout the profession as "the criterion problem," has produced tons of ratiocination, justification, and explanation. It is difficult to accept the concept that measures related to success in training for a certain job should have zero relationships with success on that job. It is improbable that zero relationships *do* obtain between those aspects of the job that are trained and the selection score. It is probable that other factors obscure the relationships if direct measures are made of global factors of job success. Evidence in support of this contention can be inferred from McReynolds' (1) finding in 1958 that successively higher aptitude levels were associated with successively higher grade in Air Force enlisted populations. This is not clean, because policies of elimination of non-productive airmen are correlated with aptitude scores.

Let us specify that objective identification of comparative levels of achievement on the job is a basic need of the science. Encouragement comes from studies based on the concept that homogeneous work on the part of all members of a criterion assessment sample enhances the probability of objective evaluation of the members of that group. Positive results from these studies were reported by Wiley (2;3) in 1964 and 1965. There is reason to believe that evaluation of proficiency at the task level may produce useful result. This, however, devolves upon future research.

What are the ingredients of the twenty-five years of personnel research under review? Casual inspection of the titles of hundreds of reports emanating in that period suggests a heterogeneity like unto an oriental bazaar. More careful screening reveals that there are certain areas which receive attention on either a continuous or a repetitive basis.

The normal human characteristic of naming that which is not understood is well known. An example or two will clarify this point. An infantry platoon is pinned down by enemy fire. The officer and sergeants are killed. Private Joe Doaks rallies the men, redeploys them, escapes the trap with them. Private Doaks demonstrated "Leadership."

A factory worker shows his supervisor how to change the design of a sub-assembly so that it involves fewer parts, is less expensive to manufacture, and is more reliable in performance. He has demonstrated "Creativity."

An individual of limited intellectual resource has spent every available minute studying and after a number of years is awarded a graduate degree by a leading university. He has demonstrated "Motivation."

These three concepts, along with others, have triggered numbers of studies over the years. Some of these studies have been generated from within, when a research worker has had an idea that suggested to him that here, at last, was the key to the problem. Others have been imposed from above, or requested by other organizations.

Consider the recurrence of some of these topics in the Laboratory program.

In 1947 and 1948 John T. Cowles and John T. Dailey, assisted by Bob Keller (4), were called upon by the Officer Candidate School (OCS) to assist in the evaluation of the candidates, in terms of dimensions like "leadership," "motivation," and "creativity." They devised a program wherein a group of officer candidates would be confronted with a problem and left to work it out under their own resources. During the time they worked on the problem, the officer candidates were observed by tactical instructors who had been counseled by the psychologists on the kinds of behavior to rate and who had been provided with forms for the collection of data.

During the experimental and training phases, everyone involved was delighted with the program. Some unrevealed leaders were discovered; some bluffers were unmasked; the judgment of relative capabilities of individuals previously held by tactical officers was supported consistently enough to please them. The system was adopted for use and then met the problem usually encountered when evaluative systems go operational.

The officer candidates' unofficial motto was "Cooperate and graduate." And they did. The word was out, and various cadets manipulated their behavior so that everyone got an equal chance to shine; the means of the ratings crept together; the variances equated; the system failed.

In 1951 Ernest Tupes (5) started collecting peer ratings in the OCS. Candidates were forced to nominate and order their fellows on various trait dimensions. These data were collected across a number of classes; they proved valid for tactical officer's ratings in the school; they were found significantly related to later Officer Effectiveness Reports (OER). The collection of such ratings is onerous, and a policy judgment came from higher headquarters that peer ratings should not be a part of the evaluative process in any officer training unit.

Factor analyses applied to trait ratings by peers in a number of different populations, and reported by Ernest Tupes and Raymond Christal (6), demonstrated a stable structure, with five common factors, across samples as diverse as kindergarten students, officer candidates, fraternity brothers, and college students.

The next cycle of this investigation involved the development of questionnaires touching the factor areas identified by the trait rating analyses by Warren T. Norman under contract at The University of Michigan (7;8;9). This work, completed in 1963, is approaching its final evaluation, in terms of validation of test scores for OERs achieved in the field.

And so it is seen that the fairly straightforward concept of evaluation of would-be officer personnel, along dimensions of behavior which are quickly recognized by name, has been treated in three distinct phases over the years between 1948 and 1966, although as yet substantive findings are not at hand. This set of circumstances highlights a major problem in the relationships between personnel research workers and operational personnel officers. Eighteen years have been spent in productive effort, and even though it now looks more promising than ever before, a method for the quick and objective evaluation of personality traits which are widely regarded as basic to desirable officer behavior is not yet available.

Impatience for results, coupled with a need for answers to pressing problems, has made administrators reluctant to take problems to personnel research workers. Not long ago a senior Air Force officer commented that by and large personnel research was worthless because the answers were always behind the problems. In a momentary sense this can be true, but in a larger view it is quite false. Consider a case which frequently rises. The general says to the psychologist, "I need the answer to Question X."

The psychologist says, "Yes sir, I can provide the answer to Question X in four years."

The general says, "My tour is three years. I cannot wait four years. I will take action Y, and you ignore the problem."

In three years and six months a new general says, "I need the answer to Question X."

The psychologist says, "Yes sir, I can provide the answer to Question X in four years," and the merry-go-round continues.

Fortunately, many persons in the system recognize this phenomenon, so that a major portion of the charter of a research unit is properly devoted to documentation of work that continues without basis in such momentary requests.

The apparent slowness of response observed in personnel research has its basis in facts which argument cannot alter. If you would educe an objective measure of some desirable trait, as for example, leadership, you must not only devise an instrument for the collection of data, but you must provide for the collection of other data which reveal the operation of the trait in question. You must give your sample of individuals upon whom data were collected the opportunity to display the variations in leadership which will demonstrate that you did, or did not, collect initial data which were predictive. This is the reason peer ratings collected in the Officer Candidate School in 1951 are not evaluated in any real sense until 1957; this is why an experimental test administered to thousands of basic airmen at Lackland in 1952 cannot be validated across one hundred technical schools until 1956; this is why a reenlistment index devised in 1960 cannot be evaluated until 1965.

There are ways to shorten these times, but in each case costs go up, and the scientific defensibility of findings goes down. We have shortened the time for validation by administering tests to airmen as they entered the technical school, accumulating a minimal acceptable sample, and correcting the derived validities for the restriction caused by administering the experimental test in an already selected group. We have collected statements of reenlistment intent simultaneously from airmen with one, two, three, and three and a half years of their first tour accomplished, and compared response patterns across the sample. These have been compared across the samples in terms of the reenlistment behavior of the first eligible sample. The succeeding samples have been added as they became eligible.

Now, let us contemplate another problem. We have discussed the objective measurement of personality traits believed basic to desirable officer behavior; we have considered the validation of a selection test for a technical school; we have contemplated the evaluation of reenlistment intent. Each of these topics is presented in a context of personnel research. At this point a number of our academic colleagues will draw their PhD cowls down tightly on their heads and snarl that the foregoing is not research and should not be dignified with such title. In-service applied research programs are but the application of recognized principles to the solution of administrative problems, and properly accomplished by technicians, not scientists.

The vicious aspect of this point is that it contains some truth; and the overt support philosophy for the in-service program has reflected it at times. There has been a period when our program documentation carried a specific indication of the weapons system toward which the research was oriented. At other times this wording has entirely disappeared, and phrases like "investigate the fundamental characteristics of thus-and-so" have taken their place. However, the consensus of senior personnel within the Personnel Research Laboratory over the years has been that the differentiation between basic and applied research is artificial, that without basic elements, the applied program is sterile and cannot last. If you force competent research personnel into a program of techniques where only momentary problems are considered, quality will evaporate, and the whole program will not be far behind.

The Personnel Research Laboratory has been fortunate that there has been a consistent sprinkling of senior officers at our intermediate and high levels of command who have recognized the imperative requirement for investigation of fundamental matters in concert with the solution of the more overt needs of the Air Force Personnel System. They have been responsible for actions such as initiating a research program in the Laboratory in the area of occupational analysis with full knowledge that nothing would come of that program during their own tenure of position. In fact, that program, initiated in 1958, produced few major findings of direct use to the Air Force until 1965, when our study of the appropriate grade for all Air Force officers reached a reportable stage (10). It is quite true that study was done in response to a direct request. That request could not have been serviced without the prior period of uninterrupted research.

It is not appropriate to refer to personnel research as if it were a univocal process, but it is beguiling to attempt to do so. Has there been any change, has there been any progress between 1941 and 1966? If so, what?

There are two factors that make such exposition difficult. The first such factor is people; the second is history. By people, it is meant that different persons in positions of authority operate from different policies, and emphasis is placed in certain areas as a function of the bias of the leader. Being in a governmental unit, we sway with the politics of the party in power; for four years we emphasize in-service research, with definite pressure to broaden our program and make it more genuinely basic science; for the next four years we minimize the in-service capability, contract for development work, and concentrate on supporting the immediate needs of weapons systems. Although such pressures have been felt, their effect within the Laboratory has not been marked. This is true because of the ponderous inertia of governmental budgetary procedures. It takes effort to induce much change in the budgetary configuration of an operation of small size, as we have been. The stability of our key staff has tended to introduce and maintain a desirable balance in the way we conduct our business.

The second factor making difficult the assessment of the evolution of the program in personnel research lies in history. If in a vacuum, with constant economic framework, with unchanging military commitments, without technological change in weapons, vehicles, computers, and communications, the evolution of the process could be observed. Since 1941 several factors external to our control have impacted on the program. First, World War II ended. A number of projects, mostly focused on development of methods of performance evaluation were inconclusively abandoned. The manning of the aviation psychology program was reduced from several hundreds to about five persons.

Starting in 1946 John T. Cowles and John T. Dailey began to assemble a peace-time research capability in military psychology. Their program grew and by 1947 monitored an aircrew testing program fostered by Reuben Baer and developed paper-and-pencil classification tests for enlisted Air Force personnel. Effort was spent identifying and elaborating measures of officer quality and broadening the coverage of aptitudes for enlisted persons by closer attention to personality factors, with data collected through biographical surveys rather than from classic personality scales. Tests were scored by machine; scores were key-punched into IBM cards; validation studies were run by computing basic summary data on IBM tabulators, then computing the Pearson correlations by hand, with desk calculators, and multiple correlations in the same way. Pedestrian and primitive by standards of today, it was a forward step from calling off scores from answer sheets to prepare scatter plots for the individual hand computation of Pearson correlations as was done early in World War II.

And then the Cold War led into the Korean conflict. Under the liberal manning policies thereby induced we found a considerable increase in Air Force psychologists. The Air Force psychology program grew rapidly and came to include research programs in operator training procedures, crew research,

maintenance techniques, methods for the use of radar observers, and a host of others. The Personnel Laboratory program grew to include research in manpower, a program in motivation and personality measurement, a program in aptitude test research, and improved support functions in experimental test administration and data analysis.

In 1957 the larger organization was curtailed, leaving the Personnel Laboratory with a program of aptitude and evaluation research, and beginning a new program of occupational analysis.

And then, one day in August of 1958, a not very prepossessing truck backed up to building T9016\* at Lackland and unloaded a few tons of hardware known as an IBM 650 Data Processing System. We didn't know it then, and we didn't recognize it for the next five years, but that was a most significant date in the Laboratory's history. We were like the cave man the day he discovered a wheel. Our load was marvelously lightened; our capacity was immeasurably increased.

At first our new computer was used as a large desk calculator. We are not unique in this respect—every organization with its first computer has gone through the cycle of automating the previous operation and of being slow to understand the true power of the computer and to apply that power economically to their need. But there had been, as there are still, men of insight in our program, and in the Air Force Personnel System. John T. Dailey, in 1948, had contracted with David Votaw (11), a mathematician at Yale, to study the personnel assignment problem, the best fit of  $N$  men in  $N$  jobs with attention to all characteristics of the men and of the jobs. Earlier Robert L. Thorndike (12) had worked on the same problems; most of the reaction to their work was amazement at the variety of symbols in the equations in their reports. In the early 1950's, John Leiman had contracted with Carl Kossack to contemplate Markov chains in the light of the flow of personnel through a system.

As we moved into the 1960's, we came of age. We modernized our computing procedures as urged by Joe Ward and Robert Bottenberg. Our analytic procedures moved five years ahead of the profession then and are five years ahead of it still, if the papers read at professional meetings are to be trusted. We have developed machine applications for the assignment models and flow models contemplated by Votaw, Kossack, and Thorndike. John Merck and Frank Ford of our staff have polished them and supervised the detail work of programming them for the computer, which was upgraded to 7040 in 1962.

More than this Joe Ward and Robert Bottenberg have gathered themselves, together with their native intellect, the analyses they encountered on the way to their doctoral degrees, the problems of a modern personnel system, and the characteristics of the high-speed electronic computer, and they have furnished the profession with a more powerful tool for the accomplishment of personnel research than was dreamed of in 1941, or 1946, or 1960 (13). Their application of the multiple linear regression model as a tool for behavioral research represents a major breakthrough in the state of the art. Precise evaluation of the role played in any personnel action by all the variables incident to that action is one manifestation of the power of the model. Data for selection of an optimal set of selection variables to meet several criterion groups is another.

As is often the case in research, minor elements of bygone programs are vital to current major efforts. During and since World War II, every bit of data or personal information on a member of the Air Force that the Laboratory happened to acquire in either a research program or from support of an operational program has been retained. Records Management types have complained of our failure to discard records more than a certain number of days or years old. Laboratory commanders have striven manfully to clean the Augean stables of our archives; "throw the junk away" has been the cry. And we did throw some few things away—after we had copied them on magnetic tape.

---

\* **Ed. Note.**—Official designation for the building which housed the Personnel Research Laboratory and housed many of its antecedent forms.

But in November of 1966, the old data are suddenly a treasure beyond price. It is possible to trace the relationships between selection scores and the actions of persons who have profoundly distinguished themselves, in either a positive or a negative way. Longitudinal studies involving selection variables and career progression become matter of course. The data bank has been augmented by the collection of information from active Air Force personnel records and by compiling them into chronological histories. Mated with aptitude, training, and career progression data, the computer can now accomplish studies of populations, not samples.

A flow model in a computer, showing numbers of persons in every category of interest, permitting the Deputy Chief of Staff for Personnel to wargame such policy changes as he may elect, cannot exist without some mystic values called transitional probabilities. It would be possible to estimate the numbers of, for example, staff sergeants in the administrative career field who have exactly seven years service who will, during the next year, become technical sergeants, remain staff sergeants, accept discharges, be killed, get married, have another dependent, switch career fields, and so forth. When the estimate is complete, it will be data of a sort, but accuracy is bound to be low. With the data bank one can develop an empirical matrix. Then one can say to the general, "Yes, sir, that is what happened."

The combination of high-speed computational facility and files in depth on personnel data permit entirely new kinds of endeavor. We have joined our files with the craft of the economist to accomplish studies which have been of financial benefit to every member of the armed services.

The facility engendered by the computer makes small work of arduous computational chores. It can accomplish the simultaneous comparison of two thousand airmen, in all possible pairs, across their recorded performance of 900 tasks in their career field. It can identify groups of airmen of homogeneous duty in a career field, even though scattered throughout the Air Force, and it can write an accurate description of the content of their jobs.

Twenty-five years have seen us move from the validation of a test battery through a process of hand-tallying of scores and the computation of correlations on a scatter plot to an automated data processing capability that can accomplish a thousand man-years of hand-done clerical work within the twinkling of an electronic circuit. The power of investigative resource has gone up millions of times with the advance in computer technology. Personnel research has broadened from the identification of aptitude variables associated with success in training to the modeling of Air Force careers and the accomplishment of demographic studies which make the chore of the personnel planner and administrator an objective science rather than a mystic art. Precision in the identification of the capabilities of persons, and the requirements of positions can now be welded to humanitarian and equitable policies for the effective manning of our military service.

#### REFERENCES

1. McReynolds, Jane. *Aptitude levels in the enlisted manpower pool of the Air Force*. WADC Technical Note 58-63 (I) and 58-63 (II), Personnel Research Laboratory, Lackland Air Force Base, Texas, 1958.
2. Wiley, L.N. *Relation of job qualification ratings to performance ratings of basic training instructors*. Technical Documentary Report 64-21, Personnel Research Laboratory, Lackland Air Force Base, Texas, 1964.
3. Wiley, L.N. Developments in analyzing airmen job performance. *Proceedings, 12th Annual Air Force Science & Engineering Symposium, Vol. II*. Aerospace medical Division, Brooks Air Force Base, Texas, 1965.

References (Continued)

4. Cowles, J.T., Dailey, J.T., & Keller, R.J. *Development of evaluation and predictive measures in the AF Officer Candidate School*. Headquarters Air Training Command, Barksdale Air Force Base, La., Research Bulletin 47-1, 1947.
5. McClure, G.E., Tupes, E.C., & Dailey, J.T. *Research on criteria of officer effectiveness*. Human Resources Research Center, Lackland Air Force Base, Texas, Research Bulletin 51-8, 1951.
6. Tupes, E.C. & Christal, R.E. *Recurrent personality factors based on trait ratings*. Technical Report 61-97, Personnel Research Laboratory, Lackland Air Force Base, Texas, 1961.
7. Norman, W.T. *Problems of response contamination in personality assessment*. Technical Note 61-43, Personnel Research Laboratory, Lackland Air Force Base, Texas, 1961.
8. Norman, W.T. *Development of self-report tests to measure personality factors identified from peer nominations*. Technical Note 61-44, Personnel Research Laboratory, Lackland Air Force Base, Texas, 1961.
9. Norman, W.T. *Validation of personality tests as measures of trait-rating factors*. Technical Documentary Report 62-4, Personnel Research Laboratory, Lackland Air Force Base, Texas, 1962.
10. Christal, R.E. *Officer grade requirements project: I. Overview*. Technical Report 65-15, Personnel Research Laboratory, Lackland Air Force Base, Texas, 1965.
11. Votaw, D.F., Jr. & Dailey, J.T. *Assignment of personnel to jobs*. Research Bulletin 52-24, Human Resources Research Center, Lackland Air Force Base, Texas, 1952.
12. Thorndike, R.L. *Research problems and techniques*. Washington: U.S. Government Printing Office, 1947, 121-127.
13. Bottenberg, R.A. & Ward, J.H., Jr. *Applied multiple linear regression*. Technical Documentary Report 63-6, Personnel Research Laboratory, Lackland Air Force Base, Texas, 1963.