

UNCLASSIFIED

AD NUMBER

ADB951150

LIMITATION CHANGES

TO:

Approved for public release; distribution is unlimited.

FROM:

Distribution authorized to U.S. Gov't. agencies and their contractors;  
Administrative/Operational Use; MAY 1952. Other requests shall be referred to Office of the Adjutant General (Army), Washington, DC 20310.

AUTHORITY

AGO per DTIC form 55

THIS PAGE IS UNCLASSIFIED

THIS REPORT HAS BEEN DELIMITED  
AND CLEARED FOR PUBLIC RELEASE  
UNDER DOD DIRECTIVE 5200.20 AND  
NO RESTRICTIONS ARE IMPOSED UPON  
ITS USE AND DISCLOSURE,

DISTRIBUTION STATEMENT A

APPROVED FOR PUBLIC RELEASE;  
DISTRIBUTION UNLIMITED,

AD B951150

# RESEARCH NOTES

UNANNOUNCED

Number 52-37 ✓

May 1952

11 May 52

12 70

DD FORM 1  
NOV 9 1979  
RECEIVED

6

SPECIFICATIONS FOR PSYCHOLOGICAL SCREENING TESTS  
FOR THE ARMED FORCES

Edward A. Rundquist

Presented to the Panel on Personnel, Committee  
on Human Resources, Research and Development  
Board, at its seventeenth meeting, 6-7 March  
1952.

~~CONFIDENTIAL~~

DDC FILE COPY

14

AGO-PRS-RESEARCH NOTE-52-37

PERSONNEL RESEARCH SECTION

10 Edward A. Rundquist

PERSONNEL RESEARCH SECTION

PR AND P BR, AGO 003 650 xll

70 11 07 140

SPECIFICATIONS FOR PSYCHOLOGICAL SCREENING TESTS  
FOR THE ARMED FORCES\*

Edward A. Rundquist

Specifications for any psychological measurement program are determined by the purposes to be achieved within the limits imposed by policy and situational conditions. In developing a measurement program, the personnel research psychologist tries to reduce the problem to the following model: determination of the criterion (the measure of the end-product that is to be predicted); developing tests that are related to this end-product; and finally combining them in the most efficient manner within the time allowable for the measurement program. Where pre-induction testing for the Armed Forces is concerned, the application of this neat formula runs into difficulty because of the changing nature of all three conditions; purpose, policies, and circumstances. An examination of how the problem must appear to high level military and civilian authority will help us arrive at some of the specifications for the pre-induction psychological measurement program.

From the military point of view, there are quite a variety of more or less related problems bearing on the pre-induction program. In the first place requirements change with the condition of mobilization -- from peace to war. During peace time, when reliance can normally be placed on voluntary enlistment as the source of manpower, and when the purpose is to develop a cadre capable of training military forces during a rapid period of expansion, a screening program will naturally emphasize the ability to absorb and pass on to others the necessary military skills. Under peace-time conditions, standards of acceptance can be kept high, the emphasis can be on general ability, the three services can compete for personnel, and in terms of motivation, work to control the desire to pass by unfair means. At least two stages of mobilization requiring different pre-induction programs may be recognized: partial and full mobilization. Two additional problems come into the picture: (1) need for qualitative allocation of manpower among the three services on some basis, and (2) the need for accepting men who cannot meet the peace-time standards -- the question "how far down can we go without decreasing the efficiency of the military" becomes the more prominent, the closer to full mobilization conditions we get. This change in requirement from peace-time to partial to full mobilization leads to a primary specification for pre-induction measurement -- we must prepare multiple programs; one program will not accomplish the different purposes.

---

\*The content of this paper reflects the opinions of the author and does not necessarily represent the viewpoint of the Department of the Army.

A second consideration from the military point of view is the necessity of every acceptee to complete the basic training of the service to which he is assigned -- boot camp for the Navy, basic training for the Army, and the indoctrination courses for the Air Force. A second specification for the pre-induction program must, therefore, be the ability to predict success in these types of training.

A third military consideration must be the wide variety of jobs for which men must be trained, many of which require technical training, and the wide variety of circumstances under which job skills, once acquired, must be applied. More than 1/3 of enlisted personnel are sent to technical schools -- to study a wide variety of technical subjects such as vehicle maintenance and cryptography. Once skills are acquired, they must be applied, not in the school setting, but under some very difficult conditions indeed. It is one thing to drive a truck in daylight over familiar terrain, with military police at the crossroads and mechanics as close as the nearest telephone. It is quite another to drive the same truck loaded with explosives, at night, without lights, over an unfamiliar road in the combat zone. Very closely related to this kind of a requirement is the variety of off-duty situations which military personnel must adjust to. Some are simple housekeeping duties, but many involve taking care of themselves among strangers both on and off the post. What all this adds up to is a three-fold requirement for pre-induction testing. Prediction of success in school training, prediction of on-the-job success, and prediction of what we may term a general adaptability.

From the military point of view, a fourth area must be taken into account -- what we will term the economics of manpower. Each individual in the Armed Forces is charged against the total authorized by Congress. Men in disciplinary barracks, their guards, men in hospitals -- some through their own negligence, personnel involved in both regular and special training -- all are charged against the total. The Armed Forces must guard against reduction in effectiveness by having to allocate too great a portion of their assets in caring for problem individuals who either cannot or will not make a real contribution to the accomplishment of assigned missions. If it takes most of the time of a competent person to care for one less competent, there is not much gained in spending the effort to bring the less competent up to minimum standards of performance. The acceptance of illiterates in the present-day Armed Forces creates a number of administrative and training problems which absorb an appreciable portion of available resources. What is gained must be balanced against this expenditure of effort. The economics of manpower lead us to two specifications for our screening program. The numbers of individuals who will be responsible for causing the Armed Forces additional problems must be taken into account. It is not simply a question of whether a particular individual can succeed after some special training, but of what the presence of given numbers will do to the efficiency of the total force. A second implication is that screening tests must be related against still other kinds of criteria: problem behavior and literacy.

As a fifth point, the military policy requiring that all processing for induction be completed within one day imposes the condition that all aspects of pre-induction processing be as brief as possible.

A final point under military considerations is a very recent and new requirement -- pre-induction psychological tests must provide one of the bases for qualitative allocation of manpower among the three services.

Before assembling the specifications for pre-induction psychological measurement, let us ponder whether there are any high-level civilian policies which might add to them. One, allocation of manpower between the military and civilian economies, might, at first glance, appear to have implications for this problem. While this is indeed a serious problem, it appears taken care of primarily by draft boards operating at the local level. Their differential deferments have attempted to gain the proper balance between national and local needs. The induction processing, therefore, does not need to take this factor into account.

There is one general consideration which, while not always made explicit, is taken into account in civilian control of induction -- the need for consistency in the operation of the induction system with our democratic tradition. This emphasizes equal treatment of the individual and equal responsibility for defense on the part of each citizen. These specifications would appear to follow: (1) measurement devices must permit equal likelihood of persons with equal ability being chosen. In other words, induction should be determined on the basis of a person's ability to learn, rather than by what he has learned. Aptitude rather than achievement tests should form the basis for the program. (2) Measurement devices must not favor any group or area. A specific example: Except when it is being measured, verbal ability which is affected to some degree by educational opportunity should be held to a minimum. More generally, we might say the tests should be either culture free or culture saturated. (3) Measurement devices should be available to detect the malingerer, if each individual is to be held to his responsibility of sharing in the nation's defenses.

A more specific, but in some ways the most far-reaching, civilian consideration is how many and how long citizens are to serve in the Armed Forces. During full mobilization, the "how many?" has the emphasis, the "how long?" being a matter of the duration of the particular emergency. During partial mobilization, both the "how many?" and the "how long?" are very critical considerations. Both the questions of type of program and standards to be applied are affected. In general, the shorter the period of service, the less time there will be for special effort on the part of the services to use marginal personnel and the higher the standards must be.

To summarize the specifications we have reviewed thus far: pre-induction psychological measurement involves multiple programs, multiple criteria, and multiple predictors. The programs are extremely sensitive to changes in either military or civilian policies or conditions and hence must be planned to permit great flexibility in operation.

The criteria that must be used in developing an induction screening test are: (1) ability to reach a minimum standard of literacy within a definite period of time; (2) success in the basic program of each service; (3) success in technical schools; (4) successful on-the-job performance under varied conditions; (5) presence or absence of problem behavior; (6) malingering behavior. In addition, the criteria for screening must take into account not only the marginal individual, but his numbers as they affect the effectiveness of units or groups. A criterion of a different kind is also needed for qualitative allocation of manpower among the services. At the present time, this is based on a proportionate division in terms of levels of general learning ability.

These criteria, along with those derived from considerations of general military and civilian conditions and policies, point to the kind of predictors which will need to be used for screening purposes. They will, first of all, be of an aptitude nature. One will surely be a pencil and paper test of general learning ability found useful in so many situations both in military and civilian life. Because of the literacy problem, a non-verbal test of this same general kind will be needed; non-verbal tests of a more specific nature will be needed to predict success in the basic training of each service. Problems of malingering will require special procedures, if not a special test. The use of a test of general learning ability for qualitative allocation makes it essential that it be constructed to cover the complete range of ability and not merely the lower levels of ability as had been the case during World War II. This test must also be particularly discriminating in the lower range where cut-off scores have been set. The entire program must be brief and capable of being administered with few and relatively untrained personnel.

With these requirements for the criteria and the predictors in mind, let us take a quick look at the past to see how the above specifications have been taken into account in the induction screening program; examine the present program with a particular eye to its validity; and look for the problems that need more emphasis in research, in order to further develop this program.

Early psychological screening was dominated by the concept that educational attainment was the essential requirement for induction. As the psychological testing movement developed -- given impetus by World War I use of tests by the Army -- pencil and paper tests were

used to some degree as a requirement for enlistment. It was not until World War II, however, that a systematic effort to develop and standardize tests for screening purposes was undertaken. The emphasis on educational attainment is shown by the nature of the first test required by the military -- a literacy test with standards to be set at the fourth grade level of attainment. The need for such a screen, of course, arose from the large number of illiterates, mentally retarded, and otherwise handicapped individuals that first the Army, and then the other services, could not deal with within the regular framework of the military establishment. In the Army, special training units were first established at the training centers and special and regular training were more or less concurrent. In 1943, the special training units were moved to reception centers and it became explicit that a goal of the induction station program was to predict graduation from STU's -- this graduation serving as a hurdle for regular basic training.

Non-verbal tests were introduced into the program before the shift of the STU's, but they were primarily designed to measure general learning ability or "intelligence" rather than to predict success in basic training. It was during 1942 that the Army began a series of studies which revealed clearly that literacy and soldier performance were largely independent abilities, each being best predicted by an almost unrelated set of tests. The validity of tests predicting graduation from special training units was of the order of .70; that for predicting success in basic training, as measured by cadre ratings, was about .50. An induction testing program was introduced which consisted of a pencil and paper test of mental ability pitched at discriminating among Grade IV and V men in terms of the AGCT. The non-English speaking were given an individual non-verbal test of general learning ability. The English speaking not qualifying in the low-level verbal test were screened thoroughly by tests designed to predict soldier performance for rejection or acceptance for STU training. This program continued until the end of the war. This quick survey of experience of World War II illustrates that, by and large, the specifications we had arrived at in our general discussion were followed during World War II. Multiple tests were used to predict multiple criteria.

The principle that pre-induction psychological measurement involves multiple programs became immediately evident at the close of World War II. The program in operation at the end of the war was a full mobilization program when every possible person was taken who showed any possibility of becoming useful to the services. As soon as the war was over, each service reverted to its own tests and its own standards until the Personnel Policy Board, established at Defense level as a result of the unification act, established the policy that the Armed Forces should construct and use a single test which could be used for two purposes: (1) screen for acceptance or rejection and (2) serve for qualitative allocation of manpower to the three services, when this might become necessary. Planning, construction and research were undertaken as a

coordinated effort by the three services in 1948. The test was put into operational use on 1 January 1950.

Most of the specifications for the AFQT have already been mentioned. Let us discuss them in more detail.

1. It was to be a test of general learning ability -- that is, an aptitude test. Content of this test should be of the type known to be effective in predicting success in a large variety of learning situations. Experience both within all three services during World War II and in civilian institutions had given ample demonstration that content like that used in the Armed Forces Classification Tests should be used: vocabulary, arithmetic reasoning and spatial items. It is well known that material of this kind is valid in predicting success in a wide variety of training situations. It is sometimes forgotten that it also has some relationship with disciplinary problems, VD rates and the like, a type of criterion of importance to the Armed Forces. For example, in one study of general prisoners received in rehabilitation centers, disciplinary barracks and federal institutions showed that the number of military offences are significantly greater among men with lower AGCT scores. Grades IV and V, although but 31.3% of the population, contributed 55.2% of the prisoner intake; in contrast, Grades I and II represented 37.3% of the population and contributed but 15.8% of the intake. Similarly, in one study it was shown that the rate of incidence of venereal disease decreased from 45 cases per 1000 total troop strength for those with AGCT scores below 70 to 2.3 per 1000 for troops scoring 130 or above. These figures are cited not as evidence for causal relationships, but simply as evidence for the fact that problems of a certain kind can be reduced through selection on a general ability test.

In saying that the new Armed Forces Qualification Test should have the type of content of the classification test used during World War II, the AGCT in effect becomes the criterion for the AFQT. One might better say the AFQT was constructed to be an alternate form of the AGCT: the two tests do correlate in the .90's.

2. The AFQT was to discriminate most accurately in the range where adjustments in the cutting score have most often been made. This range extends from the average to scores which reject from 10 to 20% of the total population of military age.

3. The AFQT was to have sufficient range to permit qualitative allocation of manpower among the three services. The test was designed to permit rough classification into 5 levels -- comparable to the five categories on the AGCT. Allocation to the three services at the present time is on a proportionate basis.

4. Other specifications for the AFQT were:
- a. It was to emphasize power rather than speed. As a means to this end, (1) the three types of items were arranged in cyclical order to permit the very slow mentally to try all types of items, and (2) time limits for the total test were set so that over 4/5 of the inductees completed the test.
  - b. The influence of education was to be reduced to a minimum. As an aid to this end, the verbal content necessary for the arithmetic and spatial items was reduced to a minimum.
  - c. Directions were to be so simple that very few would be prevented from starting the test because of inability to understand them.
  - d. The AFQT would be easy to administer.

Tests such as the AFQT are, of course, not suited to adequate assessment of the trainability of the non-English speaking and the illiterate. Much research effort went into this problem during World War II, as has been noted. So far as training criteria were concerned, these efforts were relatively successful. Test batteries were developed which could assess ability to acquire a minimum knowledge of reading, arithmetic, and the writing and understanding of English. A different battery was developed to assess ability to succeed in basic training as measured by ratings made by training cadre. These tests are available for a full mobilization program. The AFQT is in operation as the screening needed in terms of peacetime conditions. We appear to be now in what might be termed a partial mobilization condition. Both the AFQT and two of the non-verbal tests developed during World War II are in use. These together with a measure of literacy based on the easier items of AFQT form the basis for placing each selective service registrant who is not acceptable under current standards into one of four categories: Marginal literate - high mental - illiterate; illiterate - marginal mental; and rejected. Should full mobilization be necessary, this categorization would expedite the calling of these men for service.

While evidence has been presented to indicate there is considerable basis for the use in the pre-induction program of tests of the AFQT type and the non-verbal types developed in World War II, there are a number of the specifications mentioned earlier that have not been met. These furnish guides for future work.

Practically all validation work up to the present has used a training criteria of one kind or another or some very gross indices of problem behavior. The specification that screening and classification tests be related to on-the-job performance has not been studied sufficiently to determine whether it is met by present tests. It might be noted, in passing, that the number of selective factors that operate when we go beyond the training phase make the technical problem of determining the true relationship between induction and classification tests to on-the-job success extremely difficult. Work in this area is under way. More needs to be done.

Another specification that has not been met is the development of a technique for the detection of malingering. At present, there is no way of knowing exactly how widespread a problem it is. There are some indications that its significance is sometimes exaggerated. Whatever the case, a technique for detecting the malinger is badly needed. Here again, research effort is just getting started.

As a stop-gap procedure, current procedures provide a Terminal Screening Guide for an interview. This includes an individual test and specific checks of meaningful factors in the educational, social and occupational histories of men failing the AFQT. It is to be used by psychologically trained personnel. While admittedly an interim procedure, it does provide a standard procedure based on available knowledge and, as such, is an improvement over leaving the problem to the unguided ingenuity of the service personnel who happen to be assigned to examining stations.

In terms of what we have called the economics of manpower, there is the problem of the numbers of the marginally able that can be absorbed without detriment to effectiveness of training or accomplishment. What is needed here from the standpoint of screening are (1) studies of the effect on a unit's effectiveness of varying numbers of these people, and (2) facts concerning costs of the marginal risk during his total military career.

Future effort will also concern the improvement of the kinds of procedures already in use. Steps are already under way, in constructing a new form of the AFQT to include questions bearing on mechanical aptitude because of the increased mechanization of the Armed Forces. In the attempt to improve the tests which have been used to measure ability to learn by the non-English speaking and the illiterate, Dr. Rulon of Harvard University, under contract with the Army, is developing a new approach to the problem which has promise of resulting in a test which will be as free of culture bias as is possible.

It may have been noticed that no mention has been made of screening for overall adjustment to Army life, except as this goes along with general learning ability, or for locating of individuals who should be

interviewed by the examining psychiatrist. While there has been considerable work in this area, it has been omitted from this discussion on the grounds that current psychiatric theory stresses the situational elements rather than characteristics of the individual. This concept would seem to result in a coarser screen for induction on a psychiatric basis than was used in World War II. In so far as this proves to be the case, there will be less need for screening tests in this adjustment area.

Induction screening is the first step in the process of matching men to military jobs. Even during partial mobilization conditions, such as the present, the manpower problem is revealed as of more importance than ever before. During a period of full mobilization, it will be of extreme importance to promote efficiency in manpower utilization by good screening and assignment procedures. Without this, it may not be possible to utilize to the fullest extent our superior technical know-how and materiel.

Research on screening as well as on the larger problem of classification and assignment cannot be expected to yield adequate results if postponed until the all-out emergency arrives. The probable numerical inferiority of our manpower pool during an all-out emergency makes urgent the improvement of screening and classification techniques prior to its occurrence. Research directed at this improvement and at the accumulation of facts about our manpower problems in order to guide policy makers must be accomplished during peace time. Only thus can manpower utilization be accomplished in long-range terms with a minimum of the inefficiency that accrues from mis-assignment.

This discussion has not mentioned one of the very knotty problems of screening for induction: on what basis should tests be standardized? This problem will be discussed later on this program by Dr. Uhlaner.

2