EFFECTIVENESS OF SELECTION AND CLASSIFICATION TESTING

ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

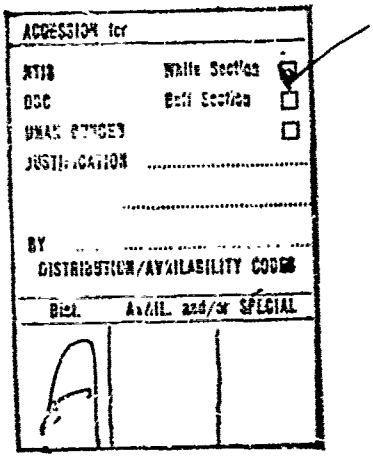
SEPTEMBER 1973

U. S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

An activity of the Chief, Research and Development

J. E. UHLANER
Technical Director

R. A. ROOTH
COL, GS
Commander



NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U. S. Army Research Institute for the Behavioral and Social Sciences, ATTN: RDMR-P, 1300 Wilson Boulevard, Arlington, Virginia 22209.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U. S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

Unclassified
Security Classification

AD 768168

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

U. S. Army Research Institute for the Behavioral and Social Sciences
Arlington, Virginia

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

3. REPORT TITLE

EFFECTIVENESS OF SELECTION AND CLASSIFICATION TESTING

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)

Milton H. Maier and Edmund F. Fuchs

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
September 1973	58	1

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO. DA RDTE Project No. 2Q062106A722	Research Report 1179
c. Training Technology & Classification-00	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Deputy Chief of Staff for Personnel, Washington, D.C. US Continental Army Cmd, Ft. Monroe, Va.

13. ABSTRACT

Army personnel managers have a continuing need to select, classify, and assign to training and jobs large numbers of men who enter the service. The Army Classification Battery (ACB) is an integral part of the assignment process; the Army Research Institute's Training Technology and Classification Work Unit Area has an ongoing research program to keep the classification battery effective and up to date. As part of the overall effort, a new ACB and aptitude area system have been developed which result in an improved system of classification for training and jobs. Technical Research Report 1177 describes these new tests, and Technical Research Note 239 evaluates them as predictors of success. The present publication addresses the value of selection and classification testing programs in relation to job training success and the suitability of the tests for subgroups of the manpower available to the Army.

The various testing programs in the Army's enlisted personnel system are described; the relationships between testing program, training content and method, and utilization on the job are probed; and the methodology is explained by which the validity of the tests is established. Analysis of measures of performance in job training programs and ratings of job performance reveals that training performance is more satisfactory than job ratings for evaluating test effectiveness. How well tests predict performance in training programs and the relation between test scores and other indexes of success are examined separately for Negroes and whites.

1-a

DD FORM 1 NOV 68 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Abilities, measure						
Aptitude area system						
Armed Forces Qualification Test (AFQT)						
Army enlisted input						
Differential classification						
Education level						
Military psychology						
*Performance, predictors of						
Psychological measurement						
*Racial group, validity for						
*Test validity						

DD FORM 1473

13. ABSTRACT continued

Selection and classification tests are effective in identifying potential failures in Army training programs and for assigning men to jobs where their potential is best used and where they can best serve the Army. Aptitude test scores are useful indicators of the proficiency and grade a man can attain, of the time required to bring a trainee to a minimum level of performance, and in identifying general categories--men eligible for officer training, for example. The tests are not only related to rate of promotion in the Army but to civilian earnings after separation from service, for both Negroes and whites. The tests are equally effective in predicting performance in job training for Negroes and whites. While Negroes make lower average scores, their training performance is correspondingly lower, comparable to that of whites with the same level of test scores.

A general criticism of tests, particularly military tests, has arisen in recent years. Analysis supports the usefulness of tests in the Army's personnel systems despite changing concepts in the Army's mission, particularly as the tests themselves are continually reexamined to meet changed requirements.

Research Report 1179

EFFECTIVENESS OF SELECTION AND CLASSIFICATION TESTING

Milton H. Maier and Edmund F. Fuchs

INDIVIDUAL TRAINING AND MANPOWER DEVELOPMENT TECHNICAL AREA
Edmund F. Fuchs, Chief

Submitted By:
E. Ralph Dusek, Director
INDIVIDUAL TRAINING & PERFORMANCE
RESEARCH LABORATORY

Approved By:
J. E. Uhlaner
TECHNICAL DIRECTOR

U. S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

Office, Chief of Research and Development
Department of the Army
1300 Wilson Boulevard, Arlington, Virginia 22209

September 1973

Army Project Number
2Q062106A722

Training Technology & Classification-00

ARI Research Reports and Technical Papers are intended for sponsors of R&D tasks and other research and military agencies. Any findings ready for implementation at the time of publication are presented in the latter part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

ARI's Training Technology & Classification Work Unit Area applies psychological measurement methods to enable the Army to make best use of the skills and aptitudes of its enlisted personnel through increasingly accurate and differentiated measures of individual potential. Research is conducted to maintain and improve the effectiveness of the Army Classification Battery and related techniques and to assess the impact of conditions that may interact with the classification tests and thus affect the basis for utilization of the enlisted input--changes in training programs and job content and environment, for example.

The present Research Report addresses the value of selection and classification testing programs in relation to job training success and the suitability of the tests for subgroups of the manpower available to the Army. The entire research work unit area is responsive to special requirements of the Deputy Chief of Staff for Personnel and the U. S. Continental Army Command, as well as to objectives of Army RDTE Project 2Q062106A722, Selection and Behavioral Evaluation, FY 1973 Work Program.

J. E. UHLANER
Technical Director

BRIEF

Requirement:

To provide a general evaluation of the Army's selection and classification testing programs in terms of their contribution to effective utilization of the manpower available to the Army.

Procedure:

The various testing programs in the Army's enlisted personnel system are described, and the relationships between testing program, training content and method, and utilization on the job are probed. A brief explanation is given of the methodology by which the effectiveness--that is, the validity--of the tests is established. Analysis of measures of performance in job training programs and ratings of performance on the job reveals that training performance is more satisfactory than job ratings for evaluating the effectiveness of selection and classification tests. How well tests predict performance in job training programs and the relationship between test scores and other indexes of success are examined separately for Negroes and whites.

Findings:

Selection and classification tests through twenty years of research and experience have demonstrated their effectiveness in identifying potential failures in Army training programs and for getting men into jobs where their potential is best utilized and they can best serve the Army. Aptitude test scores are useful indicators of the level of proficiency and grade a man can attain and of the time required to bring a trainee to a minimum level of performance. They are also useful in identifying general categories of input--men eligible for training to become officers, for example. The tests are related to rate of promotion in the Army and to civilian earnings after separation from service. Much the same order of relationship holds for Negroes and whites.

The tests have been found to be equally effective as predictors of performance in job training for Negroes and whites. While Negroes score lower on these, their training performance is correspondingly lower. Their training performance is comparable to that of whites with the same level of test scores.

Utilization of Findings:

The present report analyzes the general criticism of tests, and of military tests in particular, that have arisen in recent years. The analysis supports the usefulness of tests in the Army's personnel systems despite changing concepts in the Army's mission. Long-standing test practices in enlisted and officer programs are continually being reexamined with a view to adjusting such programs in the light of changing requirements.

CONTENTS

TABLES

FIGURES

EFFECTIVENESS OF SELECTION AND CLASSIFICATION TESTING

OVERVIEW

Each year, the U. S. Army must obtain thousands of new soldiers and prepare them to become productive workers in their units. Accomplishment of assigned missions in the Army is critical to the nation and often a matter of life and death to the individual soldier. Selection and classification testing has been used by Army personnel managers for many years to enhance the Army's capability to accomplish its mission. The effectiveness of selection and classification testing in the Army is examined in the present report. A description of the tests is given, together with an explanation of how their effectiveness is determined. Their effectiveness in predicting job training and other indexes of success is examined separately for Negroes and whites. Finally, the interrelationship of tests, training, and utilization on the job is considered.

Army personnel managers need efficient and effective means of identifying men who can profit from job training and contribute to accomplishment of unit missions; men who qualify for training must then be classified to determine the types of job in which they can profitably serve the Army.

Historical solutions to problems of selection and classification, both military and civilian, have relied on personal interviews, information about prior education and work experience, trial-and-error decisions, even random assignment. All these techniques are costly in terms of time required to arrive at a reasonable decision or in terms of errors in assignment, or both. The modern solution used by the Army and by the other armed services in this country is selection and classification testing.

The first decision the Army makes about an individual is whether he should be accessioned--that is, inducted or accepted for military service. To be accessioned a person must be mentally, physically, and morally qualified. Mental qualification is determined primarily on the basis of the Armed Forces Qualification Test (AFQT). For marginal men, supplementary evidence of trainability is required. Marginal men are those who are above the minimum level--above the bottom 10 percent of the population of young men aged 18-25--but below the top 70 percent. Most of the top 70 percent can succeed in some Army job, and they are accepted with no further mental qualification. The bottom 10 percent includes a group in which a preponderant percentage would be failures in the Army. Men in this category are therefore not accepted for service. Men in the marginal group may succeed in the Army under favorable conditions. To identify marginal men who are likely to meet training and job requirements, the Army has usually required additional evidence of trainability. This evidence is obtained from aptitude scores on the Army Qualification Battery (AQB).

For each man accessioned into Army enlisted service, the second decision is determination of the job area for which he is best qualified. The Army Classification Battery (ACB) is used to classify men according to their aptitudes or potential for success in different Army jobs.

TESTS USED IN SELECTION AND CLASSIFICATION[1]

The AFQT is a measure of general trainability composed of 100 questions, or items, equally divided among word knowledge, arithmetic reasoning, spatial perception, and knowledge of tool functions. Items similar to items in the AFQT are widely used in the civilian sector to measure potential for success in formal academic training and in selection for jobs of all types at all levels. The AFQT yields a single global assessment of potential to perform successfully in the variety of Army job training programs. The AFQT is described in Appendix A.

The Army Qualification Battery (AQB) is used as a supplementary screen to determine mental qualification by measuring the individual's strengths and weaknesses for different kinds of jobs. The current AQB assesses the person's potential in seven aptitude or job areas: Infantry, Armor-Artillery-Engineering, Electronic Maintenance, Mechanical Maintenance, General Maintenance, Clerical-Administrative, and General Technical. A Radio Code score can also be obtained for men who apply for a job in the Radio Code area. The Army Classification Battery (ACB), given to new recruits at Reception Stations, is similar to the AQB, except that the component tests are longer. Eight aptitude area scores are obtained. These are the same as those obtained from the AQB, plus a Radio Code Aptitude score.

Content of most of the tests in the AQB and the ACB is similar to that of differential aptitude batteries used in the civilian sector for vocational guidance or for job placement, or both. The tests used by the Army have been developed specifically to meet the Army's needs, and too, the General Information Test and the Classification Inventory, have subject matter unique in the Army. The tests making up the AQB and the ACB are described in Appendix B.

DEVELOPMENT AND EVALUATION OF THE TESTS THROUGH RESEARCH

The Army Research Institute for the Behavioral and Social Sciences (ARI) in the Office of the Chief of Research and Development has a continuing research program to develop and improve selection and classification tests. The research steps typically followed by ARI in developing a test and in determining its predictive accuracy--or validating it--are as follows:

---

[1] The Army Selection and classification tests described in this report were in use prior to 1 May 1973, on which date a Army Classification Battery was introduced. The principles and conclusions presented in this report are not affected by that change.

1. Definition of the problem, or the purpose for which the test or battery of tests is required.

2. Construction of the test or tests designed to measure aptitudes, interests, or other personal characteristics that predict performance in the Army.

3. Deciding how the soldier performance is to be measured. Performance may be in job training courses or on the job, and constitutes the "criterion" the tests should predict.

4. Measurement of soldier performance.

5. Analysis of the items or tests to select those that contribute to effective prediction of the criterion, and to find effective combinations of predictors. An effective test or predictor is one that shows a close correspondence to the criterion measure of performance.

Measurement of Performance

The measurement of soldier performance (Step 4) is a critical step in research to evaluate the effectiveness of a test. Evaluations of how well a man is performing on the job are not found ready-made; just as tests have to be developed to predict performance, so procedures must be developed to assess how well a man is performing in his assignment.

Ideally, assessment of performance on all relevant training and job tasks would be completely accurate. However, since job duties vary from location to location, from supervisor to supervisor, and even from time to time, no universal or absolute measure of job performance can be obtained. The best that can be accomplished is to obtain an estimate of a man's performance under a range of job conditions. The estimate is based on a sampling of what a man does on a job, not on everything that the job is likely to demand of him. A critical research problem is to find satisfactory ways of sampling the job demands to build an adequate criterion of performance against which to evaluate the tests.

Performance testing, that is, assessing how well a man does on a specified set of tasks, has appeal as a criterion of competence on the job. However, performance tests, no matter how carefully and expertly devised, can include only part of what men do and are expected to do on the job. In addition, performance testing typically covers only the minimum job requirements, and thus does not assess the critical component of how well the man can apply his skills and knowledge to new situations. The criterion should be an assessment of typical performance over an extended period of time, and performance testing ordinarily is conducted in a brief period, such as one or two days. Thus, a man may be

highly motivated to excel on the limited performance test, but his typical performance may be at a lower level. The converse may also be true: A man may get excessively nervous over the test and score below his typical level of performance.

To obtain measures of typical performance in real work environments, researchers have frequently relied on ratings of performance made by supervisors or fellow workers, or both. One assumption underlying the use of ratings is that a representative sampling of duty positions, workers, and supervisors covers most of the job demands, and thus on the whole the relevant aspects of the job are included in the criterion assessment. Another assumption is that the rater, whether supervisor or fellow worker, considers a man's performance in light of the responsibilities that are part of the job. The rating includes not only how well the man can perform the tasks but also how well he actually meets the demands; in contrast to performance tests, which assess only competence or skill in accomplishing specified tasks, ratings reflect both competence and willingness to get the job done. Ratings suffer from a shortcoming in that they are subject to great variability because there are no objective standards by which a worker's performance can be assessed. Subjective impressions play a large part in ratings, and thus ratings do not fully satisfy the scientific requirements of objectivity and standard conditions.

The criterion that is scientifically adequate and usually preferred by researchers is evaluations of performance in job training courses. Such evaluations are based on a better sampling of job demands than are performance tests, and are more objective than ratings. During job training, a man's performance is observed over an extended period of time in fairly standard situations. Since many men are being trained, each man's performance can be compared to that of a large number of others who are also learning the job. In contrast, the typical job environment has a limited number of workers at the same level, and the supervisor must base his evaluations on a limited sampling of men. In job training, objective standards of performance have been developed, and all trainees are evaluated against these standards. The assumptions underlying the use of training grades as a criterion are 1) that the curriculum builders for Army school courses, because of their extensive experience as workers and trainers, have incorporated the most important job skills and knowledge into the training programs, and 2) that all trainees go through much the same instructional materials. Grades obtained in training courses provide assessments of the men's ability to accomplish a uniform set of job tasks, and also to a degree their capacity to generalize to new situations. Granted, training grades are one step removed from performance on real-life job duties. However, because they satisfy scientific requirements for extensive sampling of tasks plus providing objective evaluations in standard situations of how well the man is doing, they are generally used by researchers as criteria.

Since almost all Army enlisted men receive extended job training before they are assigned to their first job duty position, the first decision about new accessions is to determine which job they should be trained for. Hence, the primary consideration in developing Army selection and classification tests has been to predict performance in job training rather than performance on the job. Prior to training, Army personnel managers know almost nothing about the capabilities of the newly accessioned men except their aptitude test scores, and an efficient and effective means of assessing potential is a must to enable making intelligent decisions. Once the man is assigned to training, his performance and potential can be assessed over a longer period of time, typically eight weeks, and, in most job training programs, over a wide sampling of tasks. Evaluation of how well a man is likely to perform on the job can be made more accurately during training, which is similar to the job environment. In the Army personnel system, then, selection and classification tests are fulfilling their function if they accurately predict success in job training.

Determining Effectiveness of Tests

Once the tests have been constructed and the criterion measure developed, the relationship, or correlation, between the test and criterion measure is determined. The degree of relationship shows the accuracy with which test scores can predict criterion performance, and it is ordinarily expressed as a correlation coefficient. The relationship between test and criterion performance is called "validity," and the correlation between test and criterion is referred to as the "validity coefficient." A validity coefficient of zero means no relationship and hence no prediction. A coefficient of 1.0 means that criterion performance is perfectly predictable from the test. Selection and classification tests used by the Army have predictive accuracy that falls between these extremes, usually yielding validity coefficients of .5 and .6.

The most useful interpretation of a correlation coefficient in a personnel system was developed by Hubert Brogden,[a] formerly chief scientist of what is now the Army Research Institute. Interpretation of a coefficient is in terms of the results that would obtain by using a test with no validity, at one extreme, and using a perfectly valid test, at the other extreme. A valid test improves the effectiveness of

---

[a] Brogden, H. E. On the interpretation of the correlation coefficient as a measure of predictive efficiency. Journal of Educational Psychology, 37, 65-76, 1946.

personnel decisions above those made on a chance or random basis. Say for example, that the Army accessions 1000 enlisted men on a particular day. If the men are selected without using any valid scores indicating ability to perform, then selection would be random, half the selected men would be above average, and the other half below average.

With a perfectly valid test, the true level of performance of each man can be identified beforehand. Only the 50 percent who are known to be above average in performance, for example, could be selected. Although perfect prediction is not possible in actual practice and tests with zero validity are generally not employed, these extremes provide convenient anchor points for evaluating operational tests, which fall between these two extremes.

The average performance of a randomly selected group or of a group selected on the basis of tests with zero validity is 100 on the Army standard score scale. Above average performance is indicated by scores above 100: the higher the score, the better the performance. Since an operational test has some validity, a group of men selected by the test would be above average; that is, their mean performance would be greater than 100. To determine how much greater, the other end of the performance scale must be anchored.

The high end of the performance scale is realized when a perfectly valid predictor is used. Although no test can be perfectly valid because level of performance cannot be perfectly measured, there is enough agreement about quality of performance to make perfect measurement of performance a useful concept. Assuming that level of performance can be perfectly identified, how well the selected group performs depends on what percentage of the population of eligible men has been accepted. If only men above average were desired, then men in the top 50 percent of the scores would be selected. On the Army standard score scale, this would correspond to an aptitude area score of 100 or better. On the performance scale, the top 50 percent would have an average of 116[3].

The maximum possible gain from selecting 50 percent of the men, then, is 16 points. If fewer men are selected, then the gain would be more. For example, if only the top 30 per ent are selected, the maximum gain would be 23 points. If more men are selected, the maximum gain would be less. For example, if only the bottom 10 percent were excluded, the possible gain would be 4 points, for an average of 104.

---

[3] The value 116 was statistically derived by assuming that performance is normally distributed with a standard deviation of 20 and that the bottom half of the population was not selected.

Since Army tests generally have correlation coefficients of about .6 with training performance, the gain from using them is 60 percent of the maximum. In the above examples, if the top 50 percent on the test are selected, the average performance of these selected men would be 109.6 (16 points times .60 plus 100), a gain of 9.6 points over random selection. If the top 30 percent on the test are selected, this group would have an average performance of 113.8, a gain of 13.8 points. Finally, if only the bottom 10 percent are excluded, the gain would be 2.4 points. With a correlation coefficient of .5 between test and performance, the maximum gain would be cut in half. In general, a correlation coefficient of any value can be directly translated into a percentage of the maximum gain possible as compared to random selection.

The increased level of performance can also be translated into percentage of selected men who perform above the average of the population of young men eligible for military service. If a test has no validity, 50 percent are above and 50 percent below the population average, no matter what percentage is selected. If 1000 men are examined on a test with no validity and the top 500 scorers are selected, then 250 of the selected men would be above and 250 below. If 1000 men were examined on a perfectly valid test and the top 500 were selected, then, of course all 500 would be above the population average. If a test with a validity coefficient of .6 were used to select the 500 men with the highest scores, then 350, or 70 percent, would perform above the population average. The percentage of selected men above the population average is a joint function of the validity coefficient and the proportion of the population excluded; as either the validity or proportion excluded, or both, increases, so does the percentage whose performance is above average.

PREDICTION ACHIEVED WITH ARMY SELECTION AND CLASSIFICATION TESTS

The predictive accuracy, or validity, of tests is determined, as indicated above, by the degree of relationship between test scores obtained prior to training and performance measures obtained after training or a period on the job. Correlation between AFQT score and job training performance is about .5 on the average. For the aptitude area scores, the corresponding correlation coefficients are about .6 in relevant job training courses. These predictors have been improving in validity over the years as improved tests have been developed and incorporated in the batteries. A new form of the Army Classification Battery has been developed that has an average correlation coefficient of .65 with relevant training performance. The estimated cost benefit to the Army from this increase in prediction is explained in Appendix C. The cost benefit from operational selection and classification testing programs in all services is described in Appendix D.

Aptitude measures have lower correlation with ratings of job performance, generally on the order of .2 or .3 for the more valid tests. Correlation with job performance ratings is lower than with training grades for a variety of reasons. Problems of sampling the tasks comprising a job and the different demands placed upon workers in different situations have already been discussed. Since job environments are not standardized, differences in job ratings can arise from sources other than true differences in job competence. Another reason for the lower correlation is that the overlap between the content of tests and training materials is greater than between tests and job tasks. Most selection and classification tests are paper-and-pencil tests, and training grades have a large component of scores on paper-and-pencil achievement tests given during and at the end of the training course, whereas there is little paper-and-pencil testing connected with evaluations of performance on the job. An additional and perhaps major reason for the lower correlation is that quality of work on the job is a function of many factors not present in the training situation or at least much less influential. There are no tests to predict the effect of these factors on performance at the present time. Quality of performance on the job includes knowledge of specific tasks such as knowing how to repair a carburetor, and nonintellectual factors such as motivation, attitudes, willingness to show up for work, and desire to do a good job. Both training and testing place greater emphasis on knowledge or cognitive skills than do most Army enlisted jobs. The lower prediction, then, results in part from limitations in assessing job performance and in part from different demands placed on the worker or felt by him on the job--demands not measured by the tests.

Prediction of Performance in Job Training for Negroes and Whites

A question of considerable social importance concerns the effectiveness of Army selection and classification tests for Negroes and whites. Tests have come under severe criticism and legal attack in the 1960's on the grounds that they are discriminatory against Negroes and other minority groups. Questions about fairness, equity, or discrimination in personnel selection, training, and utilization on the job involve issues that extend beyond scientific research. What scientific research can do is answer questions about how well Negroes and whites perform on tests and on the criterion and about the relationship between test scores and later performance for the two groups. Findings from scientific research can be considered when making policy decisions about fairness in testing Scientific research, however, does not by itself determine any particular policy.

To answer the scientific question about the effectiveness of Army tests for Negroes and whites, test scores and job training performance were analyzed for over 12,000 white and 1700 Negro enlisted men. The men were tested before their job training began and then followed up during training to determine how well they did.

Effectiveness of the tests was studied in six job areas: combat arms (including infantry, armor, and combat engineering, but not field artillery), electronics repair, motor maintenance, general maintenance, clerical-administrative, and general technical. Mean or average level of performance on the relevant aptitude tests and in job training was computed for each racial group in each job area. The predictive accuracy, or validity coefficient, was also computed for each racial group. These results are statistically adjusted to provide estimates of results that would be obtained for random samples of Negroes and whites. The results can therefore be generalized to the population of all young Negro and white males eligible for military service. The predictive accuracy is shown by graphs relating test scores and training performance (Figure 1). Table 1 gives the data from which the graphs were plotted.

The results for electronics repair are described in some detail, since they are typical of what was found. The lines relating aptitude scores and training performance for Negroes and whites are essentially the same; this similarity means that the tests used to predict success in electronics repair are about equally effective for the two groups. Since the differences are small, the conclusion is that training performance for both groups is at about the level expected from the test scores. For example, both Negroes and whites with scores of 100, which is the population average, are expected to show average performance in training. Low-scoring Negroes and whites are expected to do less well in training, and the lower the test score the lower the expected performance. In terms of predictive effectiveness to identify men who can succeed in training for electronics repair jobs, the Army selection and classification tests are equally valid for Negroes and whites. In three other job areas, motor maintenance, clerical-administrative, and general technical, the results were similar to those for the electronics repair area.

In the general maintenance area, the line for whites is somewhat higher than that for Negroes. In this area, Negroes did slightly less well in training than did whites with the same test scores. For example, Negroes with a General Maintenance aptitude area score of 100 have a predicted training performance level of 96, while whites with the same aptitude area score have a predicted training performance level of 100. In other words, there is a slight tendency for the aptitude tests to favor Negroes as compared to whites for purposes of predicting training performance in general maintenance jobs. Since the absolute differences in level of predicted performance are small, and since the results for

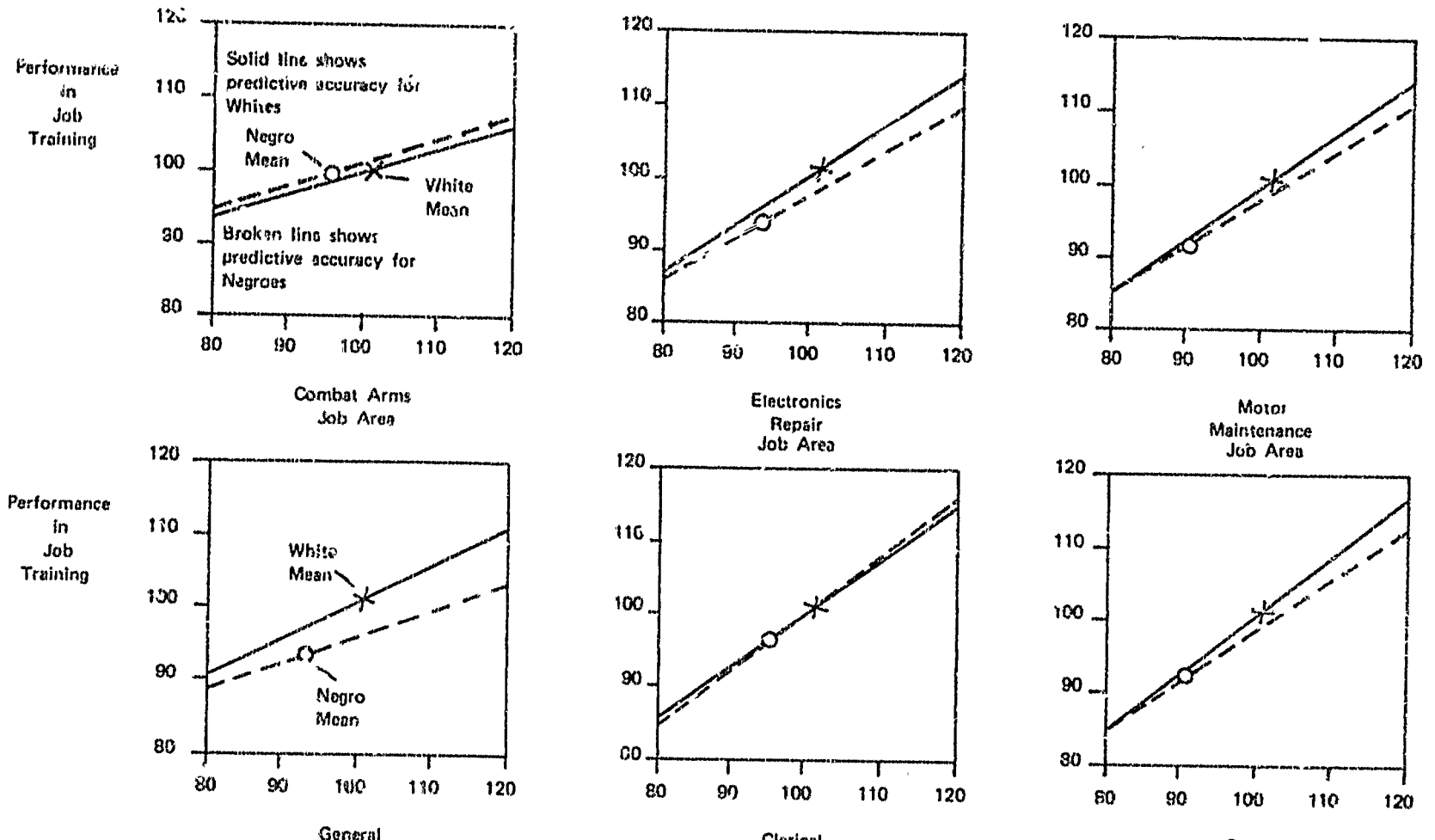


Figure 1. Predictive accuracy of classification tests for predicting performance in job training in specific job areas for Negro and White men.

Table 1

STATISTICAL INDEXES PLOTTED IN FIGURE 1

Job Area	Correlation Coefficient		Standard Deviation			
			Job Training		Classification Test	
	Whites	Negroes	Whites	Negroes	Whites	Negroes
Combat Arms	.32	.35	21.4	19.7	20.0	20.0
Electronics Repair	.62	.53	23.6	23.2	20.0	20.0
Motor Maintenance	.65	.59	23.2	21.6	20.0	20.0
General Maintenance	.47	.32	22.0	22.4	20.0	20.0
Clerical-Administrative	.64	.67	23.7	24.4	20.0	20.0
General Technical	.67	.57	23.4	23.9	20.0	20.0

the other job areas do not support the difference found for general maintenance, no great significance can be attached to the results for this one job area.

In the combat arms training courses, the line relating training and test performance for Negroes is slightly above that for whites. In these job training courses, then, Negroes perform slightly better than expected from their test scores. As for the other job areas, the differences are slight, and the conclusion remains that the tests work equally well for the two groups.

The mean level of test scores and training performance is also shown in Figure 1. X indicates the means for whites and O the means for Negroes. Training performance and test scores were set to have means of 100 for the combined racial samples. Means were then computed separately for Negroes and whites. In all job areas except combat arms, the mean for whites is to the right and above that for Negroes. Whites score higher both on predictor tests and in training. The difference in predictor means is about 8 points in the clerical-administrative and combat arms areas and extends up to 10 or 11 points in motor maintenance and general technical. The differences in training performance means are smaller: a maximum of 9 points in motor maintenance and general technical; 5 to 8 points in clerical-administrative, general maintenance, and electronics repair; and virtually no difference in combat arms.

The smaller difference between the groups in training performance is just what is to be expected from statistical theory. Any group below average on a predictor that is less than perfectly valid will score relatively higher on the criterion, but still below average. The technical label for this shift in means is called "regression toward the mean." Since most Negroes score below the mean of whites on the predictor, their training performance is relatively better. Similarly, the training performance of a group of whites with the same set of predictor scores is better than their predicted performance.

To summarize the statistical results, the evidence indicates that in terms of predicting training performance, Army selection and classification test scores have the same meaning for Negroes and whites. To be sure, Negroes score lower on the predictor tests, but their training performance is correspondingly lower. These results are consistent with those found by the Air Force where job training was predicted by tests about equally well for Negroes and whites and with results in many civilian colleges where tests were about equally effective as predictors of college success for Negroes and whites.

The fact that the means for Negroes on tests are well below those of whites has led to suggestions that lower prerequisites for job training or minimum qualifying scores be established for Negroes. An

alternative way of phrasing this proposal is to establish separate norms for each racial group, and then select men from each group separately. For example, the top 50 percent of each could be eligible for a particular job training program. While this proposal would insure that a certain number of Negroes would be accepted for job training, it would by no means insure that many of them would complete training successfully. In fact, the evidence indicates that the failure rate for Negroes would be increased under the lowered prerequisites proposed. With the present prerequisites, their training mean performance is well below average, and to admit more men with lower test scores would only serve to lower mean performance even further.

It should be stressed that these results were obtained in a particular context. Training performance of Negroes could perhaps be changed by changing other components of the personnel system. Training objectives could be modified by fractionating existing jobs; training methods and materials could be adapted to accommodate lower level students. If these changes were made, then more Negroes could be accepted with a greater likelihood of success. The point to be made is that modifying only the test score prerequisites for Negroes while leaving training programs intact would serve to increase their failure rates. Test scores are only one component in the total process of selecting, training, and utilizing a man on the job. An effective system requires that these three functions complement one another, and if one is adjusted the other two must be brought into line.

A final word about equity or discrimination in testing. Strictly speaking, discrimination is not a scientific issue; rather, it is a legal and moral issue. The scientist can provide evidence on which intelligent decisions can be based. The scientific evidence presented here suggests that the issue of equal employment opportunity is more complex than merely adjusting test scores to obtain equal means.

Other Evidence of Importance of AFQT Scores

In addition to predicting success in job training, selection and classification tests have been found to be related to other measures of success in life--in fact, to a variety of indexes of social and job status. For example, surveys of Army personnel consistently show that AFQT score bears a strong relationship to enlisted grade. Men scoring higher on the test get promoted faster and they tend to reach higher grades. In a survey of men separated from the Army after from 19 to 25 months of service, the percentage of men at grade E-5 and above for different levels of AFQT score and controlled by race and level of education were as follows:

Race	AFQT Score	Non High School Graduate	High School Graduate
White	21 - 30	26%	34%
	71 - 80	33%	50%
	91 -100	45%	59%
Negro	21 - 30	21%	30%
	71 - 80	31%	40%
	91 -100	33%	58%

The percentage of E-5's and higher increased steadily with AFQT score for both Negroes and whites and for both levels of education.

The Department of Defense analyzed the post-service earnings of enlisted men ten months after separation from the Army. For both Negroes and whites, mean income showed a steady rise with higher AFQT score. For example, for whites in the AFQT score range 20-29 the mean assessed income was about $6500 compared to $7400 for whites in the 80-89 score range. For Negroes, the comparable figures were $6200 and $7000. Unemployment figures showed a similar trend. For whites scoring 10 to 30 on the AFQT, 7.3 percent were unemployed ten months after separation as compared to 5.2 percent of men scoring 93 to 100. For Negroes the figures were 12.0 percent and 2.8 percent, respectively.

The Army in late 1971 set new policy on the granting of moral waivers. Under the revised policy, a person with a felony conviction and who dropped out of high school or scored below 31 on the AFQT, or both, is not ordinarily granted a moral waiver. Persons with a felony conviction who are high school graduates and score above 30 on the AFQT can be granted waivers. The decision to use AFQT score as part of the basis for granting moral waivers is supported by years of research and experience in which the consistent result is that men scoring low on AFQT have a greater tendency to become disciplinary cases.

Systems Analysis Of Testing

Army selection and classification tests have been shown through years of research and operational use to fulfill their intended purpose--to obtain optimal performance from available manpower resources. The tests have consistently been found to be related to valid measures of potential to perform, and thus have enabled more intelligent personnel decisions.

Effective management of manpower resources always occurs in a specific context or system. Thus, the validity of the tests has been established for a particular job training program, which in turn was designed to qualify workers for a specified job structure. These three components of the personnel system--selection, training, and utilization on the job--are critical dimensions; as one changes, the others are also affected.

Put another way, a test has validity only for certain criteria and not for others. Selection and classification in the Army traditionally have been used to identify men with the greatest potential to succeed and generally to reject men with the least potential.

In the late 1960's, the Department of Defense formulated policy on human goals that could have implications for the way tests are used in the future. One of the goals was "to contribute to the improvement of our society, including its disadvantaged members, by greater utilization of our human and physical resources while maintaining full effectiveness in the performance of our primary mission." This goal suggests that personnel management in the military services must consider not only effective performance, but also the impact of personnel decisions on all elements of American society. In a society where training and promotion were based on merit, tests served well. In a new context where greater utilization of manpower resources from all walks of life is a goal, background characteristics may need to be considered in making personnel decisions, and order of merit as identified by test scores may no longer be the controlling factor.

An important change in the effectiveness of tests has been found when they are used to predict success in the new type of Army job training called "performance-referenced training." In the new training, a limited number of tasks are specified as essential to a job. All men assigned to this type of training are trained to perform at a minimal level of competence--the mastery level--and are then graduated as qualified workers. Differences in performance above mastery level are not considered in the evaluation of training outcomes. In contrast, traditional job training has placed a high ceiling on what a man should know after training. In a sense, the traditional style of instruction was designed to develop in each person the maximal level of skill and knowledge commensurate with his aptitudes rather than being concerned mainly with acquisition of minimal skills. Validity of the Army test was established for the traditional type of training.

In the new type of training program, tests would be expected to have little or no validity in predicting level of performance. Tests can have validity only if there are differences in levels of performance. Since differences in performance are not assessed in performance-referenced training, the criterion of success is irrelevant for determining the effectiveness of aptitude tests. Thus, the fact that completion of performance-referenced training is unrelated to test scores in no way reflects on the validity of the tests. Since there are differences in time required to attain mastery level of performance, the tests remain valid for predicting which men can learn the material the fastest.

The new training program could also require reconsideration of how men are utilized on the job. As was discussed earlier, traditional training courses serve the function of additional selection of men through extensive observation and evaluation of performance during the

period of instruction. If everyone completes training satisfactorily, then this additional selection would need to be performed subsequently on the job. Since at present the job environment has few procedures for performing selection, procedures would need to be developed to identify men having difficulty in adjusting or coping with the tasks encountered on the job and to effect transfers to different jobs when necessary.

Even though tests would lose their validity for predicting success in performance-referenced training, they would still be valid for predicting performance during the initial learning phases on the job. If the men are taught skills that involve reasoning and memory, aptitude tests will predict both level of performance and ease of acquiring the skills. Jobs differ in the level and complexity of demands imposed on the incumbents, and the differences between individuals in ability to perform will emerge. When the criterion situation allows differences between individuals to surface, then the differences in aptitude test scores will be related to level of performance.

The interrelationship of job level, training and mental ability in the current Army personnel system is shown in Figure 2. Four job levels are considered--commissioned officer, noncommissioned officer, skilled journeyman, and unskilled worker. The commissioned officer level is the most complex and requires the highest level of mental ability, ordinarily in the top third of the population. Technical skills acquired through experience ordinarily cannot be substituted for high mental ability, especially at senior officer level. At noncommissioned officer level, most men are in the top half of the population in mental ability. In some cases, noncommissioned officers with lower mental ability than is ordinarily required have been successful. Extensive training or experience can sometimes compensate for lower ability, especially in less technical areas, such as combat arms. At the entry levels of most Army enlisted jobs, rigorous training can prepare most men to perform adequately. However future leaders and supervisors in the technical areas are ordinarily drawn from among men with above average general ability.

The usefulness of tests for men who are average or higher in ability is generally well accepted. For example, selection test are widely used to pick college students and commissioned officers. There is some concern, however, that the tests are not useful for men with low ability. The criticism is often based on the fact that most Army tests are of the paper-and-pencil variety. The concern is that less able men lack familiarity with the verbalization in the tests and perhaps are not even accustomed

Job Level	Job Requirements	Usual Training or Educational Requirements	Accepted Aptitude Requirements
Commissioned Officer	Management of people and things	College graduation or equivalent	Top third of population in general mental ability
	Organization and direction of resources to accomplish mission	Leadership	High verbal and reasoning skills
	Accurate and timely decision making	Managerial and administrative skills	
	Versatility to cope with constantly changing environment	Technical proficiency and job experience ordinarily is not an acceptable substitute for high general mental ability	
Non-Commissioned Officer	Supervision of workers	High school graduation or equivalent	Top half of population in general mental ability
	Interpretation and implementation of regulations, directives policy statements and orders	Supervisory and administrative skills	Top half of population in specific aptitude job area
	Detailed knowledge of technical aspects of job area	Technical competence in job area	
	Ability to adjust to new situations	Experience and training on the job can sometimes be used as substitute for general ability; high specific aptitude ordinarily is essential	

Figure 2. Relationship among job level, and job training, and aptitude requirements in the Army

Job Level	Job Requirements	Usual Training or Educational Requirements	Accepted Aptitude Requirements
Journeyman in skilled job	Proficiency in job tasks Reading assimilation of technical regulations and manuals Comprehension of verbal orders and instructions	Training to perform job tasks Knowledge of underlying theory to enable adaptability to work on new equipment Literacy at high school level Experience and training can be used as substitute for general ability and for specific aptitude when work is in standardized situations	Top two-thirds of population in general mental ability Top half of population on specific aptitude for job area
Unskilled Job	Application of common sense to carry out simple instructions Work in standardized situations	Functional literacy to understand simple verbal and written instructions Training ordinarily received on the job or in short formal courses Experience and training can be used on substitutes for aptitude, provided functional literacy level is attained	Top three quarters on general mental ability or in top nine-tenths if supplemented by additional evidence of trainability such as specific aptitudes

Figure 2. (continued)

to using pencils. However, the statistical evidence presented earlier about the validity of aptitude tests for Negroes and whites shows that paper-and-pencil tests are effective predictors in traditional job training programs for men of low ability. For these men, there is consistency of performance between test and training situations. Such consistency is the mark of validity. Whatever factors operate to hold down test performance for these men also depress training performance.

In summary, Army tests have been in use for many years. Initially, there was resistance to their operational use. Then, with demonstrated results in the form of improved selection and assignment, a general acceptance grew up. Individual points of criticism were made and these have at times resulted in improvements. In recent years, tests have come under increased criticism both in and out of the Army. The criticisms have prompted the testing research community to reexamine many long-standing beliefs and practices. The likely outcome is that tests will continue to be needed and to be used, but the circumstances of their use and in some cases the tests themselves will adjust to changing concepts of their role.

APPENDIXES

APPENDIX A

THE ARMED FORCES QUALIFICATION TEST

(AFQT)

Information Pamphlet

PT 0502R

ARMED FORCES QUALIFICATION TEST
INFORMATION PAMPHLET

PURPOSE OF THIS PAMPHLET

This pamphlet has been prepared primarily to give general information about the AFQT--the test you will take at the Armed Forces Examining Station. AFQT is an abbreviation for Armed Forces Qualification Test, which is the mental ability test all men have to take before going into one of the Armed Services.

It is hoped that after reading this pamphlet you will understand better the purpose and make-up of the AFQT and thus be more at ease when you take the test.

WHY AN AFQT?

The AFQT has been compared to a screen--a mental screen. It sorts men in terms of their ability level and throws back or rejects those who do not come up to the level required by the Armed Services. Why should all men be required to take--and pass--such a screen before they can be accepted for service? Isn't a physical examination enough? Can't anyone who is healthy be a soldier? The answer is "No." It takes more than physical ability to learn the hundreds of details in basic training, and in advanced individual training, and to learn to do the many tasks that make up the job of today's soldier. The ability to learn these things is what the AFQT measures.

WHAT IS THE AFQT?

The AFQT has been called a test of "military trainability" for the logical reason that men who do well on it can absorb military training without difficulty and learn to do a good job in the Armed Forces. A man does not have to be a genius to pass the AFQT--or to do well on it. It is not an "I.Q." test. It is not an "education" test, although those who have done well in school usually make good scores on the AFQT. But there are also men who did not go very far in school who do well on the AFQT.

DO YOUR BEST

Congress has set a minimum score for qualifying on the AFQT. But just barely passing won't do much for a man other than get him a uniform. In order to be selected for training in specialist schools, or for Officer Candidate School, or for promotion to the more interesting and better paying jobs, a man needs to demonstrate more ability than just

being able to pass. Whether you want to make progress in the military or develop skills which will be useful to you later in civilian life--you should do your best and make the highest score you can on the AFQT. And you will have only one chance to do this. Men are not given the AFQT a second time just because they did poorly the first time. So do the very best you can on it.

TAKING THE AFQT

When you go to the Armed Forces Examining Station to take the AFQT, you will be given a test booklet containing several practice questions and 100 test questions. You will also be given a separate answer sheet where you will mark your answers to the test questions. A special pencil with an eraser will be provided along with blank paper where you can do any figuring you want.

You will receive complete instructions as to what to do in taking the test. Some of these instructions are printed in the test booklet; all are read aloud by the examiner conducting the test. You will be told how to mark your answers on the answer sheet in the proper spaces. If you make a mistake and mark the wrong answer, you can erase it and mark a different one.

After you have been told what to do to answer the questions, you practice by answering several sample questions. You are told the correct answers for these practice questions. Finally, you are given an opportunity to ask questions of the examiner if you do not understand what is expected of you.

THE TEST AND THE ANSWER SHEET

You will find four types of questions in the AFQT. There are questions about the meaning of words, questions about arithmetic, questions about tools, and questions about boxes made by folding pieces of cardboard. Sample questions of each of these types are shown on the following pages. There are 25 questions of each type or a total of 100 questions in the test. Fifty minutes are allowed for taking the test. Not everyone is able to finish the test in this time, but you can make a good score even without finishing it.

Each question in the AFQT is followed by four possible answers labeled A, B, C, and D. Only one of these is right. Your job is to decide which of the answers is right and then mark the answer sheet (<u>not</u> the test booklet) to show your choice.

There are 3 main parts to the answer sheet. One part provides space for your name, the date, and other such information. A second part contains a space with heavy lines around it labeled "Practice Questions."

This is for the answers to the practice questions which everyone takes before beginning the 100 questions that count. The rest of the answer sheet--the main part--is numbered from 1 to 100 to provide spaces for the answer to each of the 100 test questions.

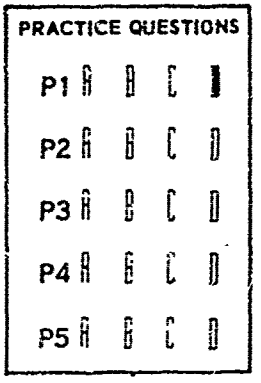
HOW TO ANSWER THE QUESTIONS

After each question in the test booklet are four possible answers next to the letters A, B, C, and D. On the answer sheet after each question number are the four letters, A, B, C, D, printed in tall thin type. You blacken out one of these letters on the answer sheet to indicate your answer to each question. Here is the first Practice Question in the test:

P1. It was a <u>small</u> table.
A) sturdy
B) round
C) cheap
D) little

The question here is: Which of the four words--A) sturdy, B) round, C) cheap, D) little--means the same as the word with the line under it, the word <u>small</u>?

The D answer, little, means the same as small, so on the answer sheet in the box labeled Practice Questions and on line number P1, you would blacken out the letter D. The answer sheet would then look like this:

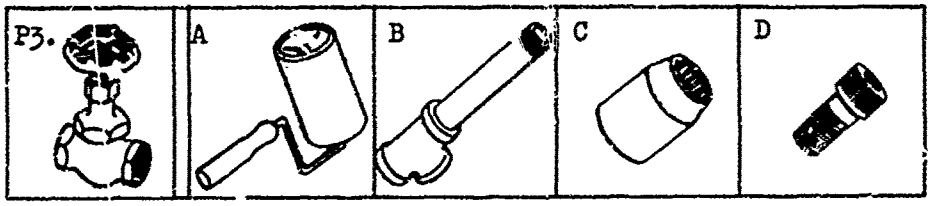


The next practice question, P2, is about arithmetic.

P2. One box of nails costs 30¢
How much do 3 boxes cost?
A) 33¢
B) 60¢
C) 90¢
D) 99¢

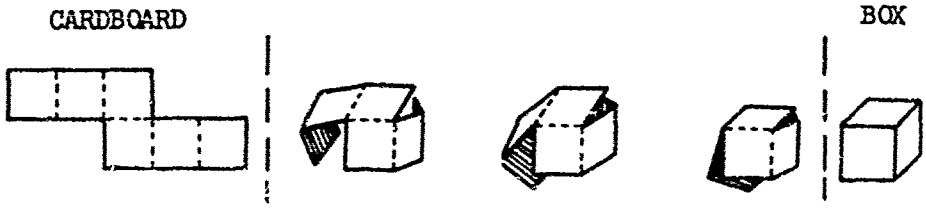
The right answer is the C answer, 90¢, so on the answer sheet the C on line P2 would be blackened out.

The third practice question is a picture question about tools. Picture questions go across the page. The question is always about the first picture in the row. You have to decide which one of the other 4 pictures in the same row goes best with the first picture.



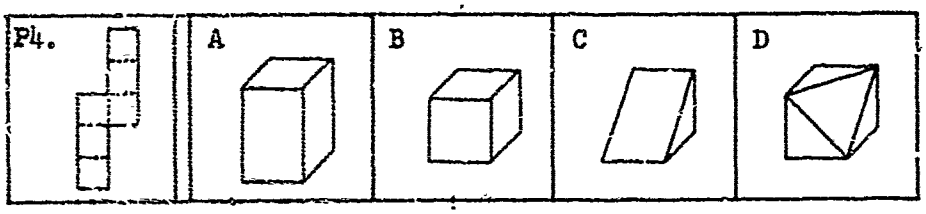
The first picture above shows a valve used to turn water on and off. The picture that goes best with the valve is the B picture, the piece of water pipe, so the B on the P3 line of the answer sheet should be blackened out.

There are two main kinds of questions about boxes. In one the job is to find the box that can be made by folding a piece of cardboard. In the other kind the question is what the piece of cardboard would look like if the box were unfolded. The row of pictures right below shows what is meant by folding a piece of cardboard to make a box.



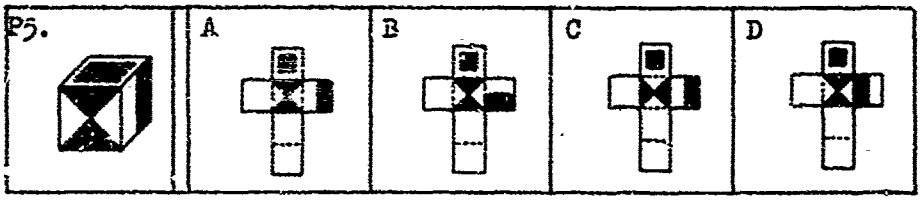
The dotted lines show where folds are to be made. The last picture shows the box that has been made by the folding.

Now look at practice question P4 below. The first picture shows the piece of cardboard that is to be folded. You are to find the drawing of the box A, B, C, or D that would be made by folding this cardboard.



The correct answer is B. So, on the answer sheet, the letter B on the P4 line would be blackened out.

In the next question, P5, if the box in the first drawing were unfolded it would look like one of the four cardboard patterns shown in the other pictures in the row. Which one?



The correct answer is A.

If you marked all the practice questions correctly, the Practice Box on the answer sheet would look like this:

PRACTICE QUESTIONS				
P1	A	B	C	D
P2	A	B	C	D
P3	A	B	C	D
P4	A	B	C	D
P5	A	B	C	D

A "SAMPLE" AFQT

The instructions and practice questions above are the same as those given to all men taking the AFQT at the Armed Forces Examining Stations.

The next 4 pages contain a sample of 16 questions like those in the AFQT--four questions of each of the four types in the current test. Try them now for practice. When you answer them use the answer space in the middle of this page. You will probably notice that the questions get harder as you go from the first to the fourth in each type. This is the way the AFQT is arranged--easy questions first then harder as you go along. About nine out of ten met get the first question right. About one out of three of those taking the AFQT are able to get the last one right. How many can you answer correctly?

Mark your answers here:

1	A	B	C	D	5	A	B	C	D	9	A	B	C	D	13	A	B	C	D
2	A	B	C	D	6	A	B	C	D	10	A	B	C	D	14	A	B	C	D
3	A	B	C	D	7	A	B	C	D	11	A	B	C	D	15	A	B	C	D
4	A	B	C	D	8	A	B	C	D	12	A	B	C	D	16	A	B	C	D

The correct answers to all of the 16 questions--just in case you want to check--are on the last page of this pamphlet.

SAMPLE AFQT QUESTIONS

1. The boys <u>discovered</u> the cave.
   A) searched
   B) found
   C) enlarged
   D) entered

2. There will be <u>variable</u> winds.
   A) shifting
   B) chilling
   C) steady
   D) mild

3. He is a man of great <u>vigor</u>.
   A) wickedness
   B) strength
   C) reputation
   D) wisdom

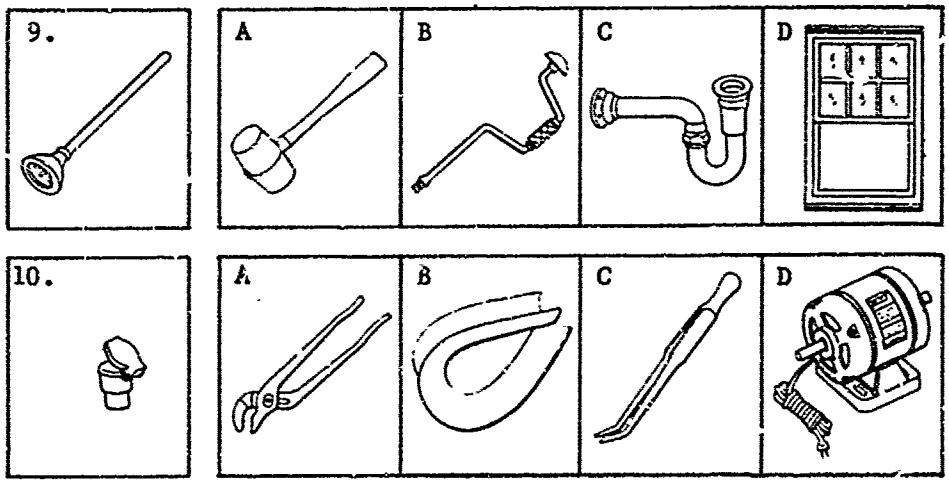
4. The prognostication proved correct.
   A) solution
   B) measurement
   C) identification
   D) prediction

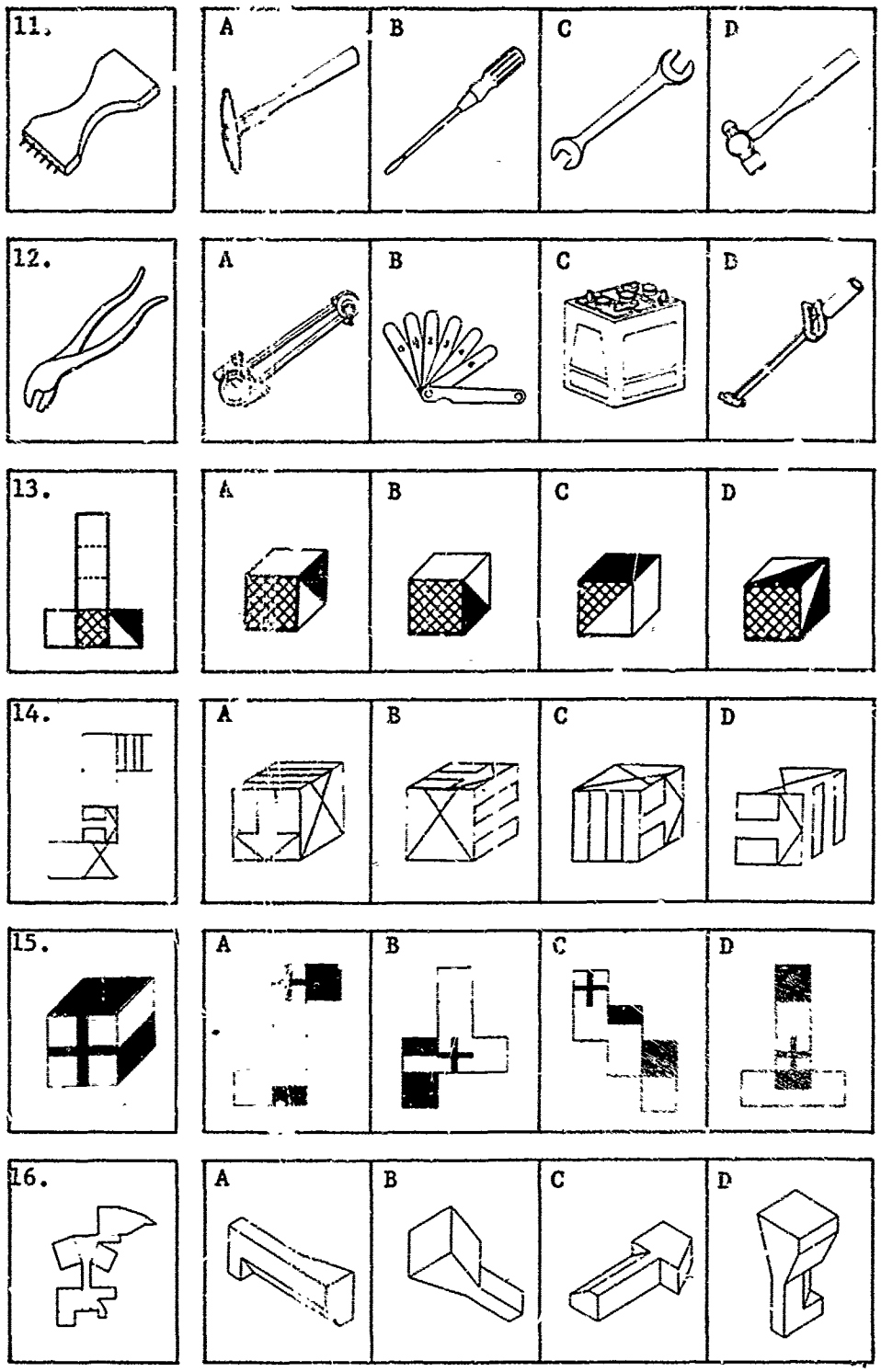
5. Bob wants to buy a wagon. He has 5 dollars and needs 5 dollars more. How much does the wagon cost?
   A) $10
   B) $15
   C) $25
   D) $55

6. If 12 men are needed to run 4 machines how many men are needed to run 20 machines?
   A) 24
   B) 48
   C) 60
   D) 80

7. During one year the fruit crop was 500 bushels. The next year the crop increased 102 percent. How many bushels were produced in both years?
   A) 510
   B) 602
   C) 1010
   D) 1510

8. If John can stack 2/3 of a cord of wood in 1 hour how long will it take him to stack 4 cords of wood?
   A) 6 hrs
   B) 5 hrs
   C) 4 hrs 20 mins
   D) 2 hrs 40 mins




11.
A
B
C
D
12.
A
B
C
D
13.
A
B
C
D
14.
A
B
C
D
15.
A
B
C
D
16.
A
B
C
D

KEY TO 16 QUESTIONS IN PRACTICE AFQT

(The correct answer to each question has been blackened out)

1	A	**B**	C	D	5	**B**	C	D	9	A	B	**C**	D	13	**B**	C	D
2	**B**	C	D	6	A	**B**	D	10	A	B	C	**D**	14	A	**B**	C	D
3	A	**B**	C	D	7	A	B	**C**	11	**B**	C	D	15	A	**B**	C	D
4	A	B	C	**D**	8	**B**	C	D	12	A	B	**C**	D	16	A	B	C	**D**

A FINAL WORD

If you have read each page of this pamphlet, and have answered the sample questions--or as many of them as you could--you should be well prepared for taking the AFQT at the Armed Forces Examining Station. You may not be able to answer every question correctly--very few people are--but you will be able to understand what is expected of you. And you should feel at ease because you will be taking a test that you are familiar with--one on which you will be able to make a score that will do full credit to your ability.

APPENDIX B ACB AND AQB SAMPLE QUESTIONS

ACB and AQB SAMPLE QUESTIONS

The following pages contain sample questions for tests composing the Army Classification Battery (ACB) and the Army Qualification Battery (AQB). Included within the AQB are four tests which compose the Armed Forces Qualification Test (AFQT). The four AQB tests used for the AFQT score are Verbal (VE), Arithmetic Reasoning (AR), Shop Mechanics (SM), and Pattern Analysis (PA).

The AFQT is the basic screening instrument used for both enlistment and induction. The AQB is used as an additional screening instrument for some inductees and as a qualifying instrument for men who want to enlist with the assurance of training in a particular occupational area or initial assignment to a particular overseas command. The ACB is a set of tests paralleling those of the AQB, but with more questions that discriminate among the higher level men. The ACB is used to identify the areas in which a man is better suited so that he may be assigned to a job where he is more likely to succeed.

The sample questions are representative of items in the AQB and ACB. The batteries represented and the number of items in each test are indicated at the top of each page. Organization of the sample questions by test and battery is shown below.

Test	Battery	Test	Battery
VE	ACB,AQB,AFQT	AI	ACB,AQB
AR	ACB,AQB,AFQT	ELI	ACB
SM	ACB	ELI	AQB
SM	AQB,AFQT	CI	ACB,AQB
PA	ACB	GIT	ACB,AQB
PA	AQB,AFQT	ACS	ACB,AQB
MA	ACB,AQB	ARC	ACB

VERBAL TEST (VE)
SAMPLE QUESTIONS

ACB: 50 items AFQT and AQB: 25 items

Each item requires the examinee to select the correct synonym for the underlined word in a short sentence.

1. The boys <u>discovered</u> the cave.

   A) searched
   B) found
   C) enlarged
   D) entered

2. There will be <u>variable</u> winds.

   A) shifting
   B) chilling
   C) steady
   D) mild

3. He is a man of great <u>vigor</u>.

   A) wickedness
   B) strength
   C) reputation
   D) wisdom

4. The <u>prognostication</u> proved correct.

   A) solution
   B) measurement
   C) identification
   D) prediction

ARITHMETIC REASONING TEST (AR)
SAMPLE QUESTIONS

ACB: 40 items AFQT and AQB: 25 items

Each item is a reasoning problem involving application of arithmetic processes.

1. Bob wants to buy a wagon. He has 5 dollars and needs 5 dollars more. How much does the wagon cost?

   A) $10
   B) $15
   C) $25
   D) $55

2. If 12 men are needed to run 4 machines, how many men are needed to run 20 machines?

   A) 24
   B) 48
   C) 60
   D) 80

3. During one year, the fruit crop was 500 bushels. The next year, the crop increased 102 percent. How many bushels were produced in both years?

   A) 510
   B) 602
   C) 1010
   D) 1510

4. If John can stack 2/3 of a cord of wood in 1 hour, how long will it take him to stack 4 cords of wood?

   A) 8 hrs
   B) 5 hrs
   C) 4 hrs 20 mins
   D) 2 hrs 40 mins

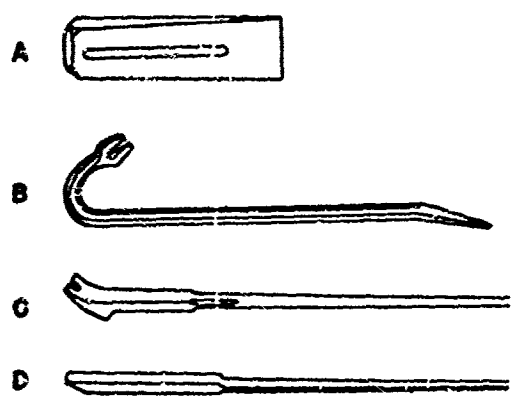
ACB

SHOP MECHANICS TEST (SM): 40 items

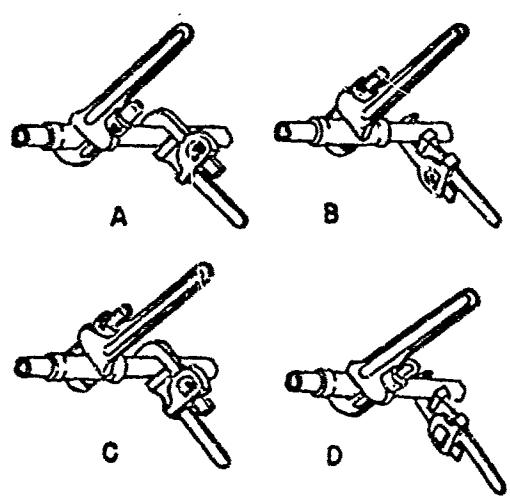
SAMPLE QUESTIONS

Questions are directed toward drawings of tools or illustrations of mechanical principles involved in shop work.

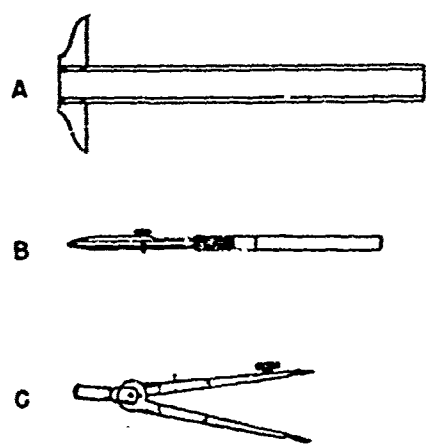
1. The tool normally used to pry boards loose and to pull nails from construction forms is



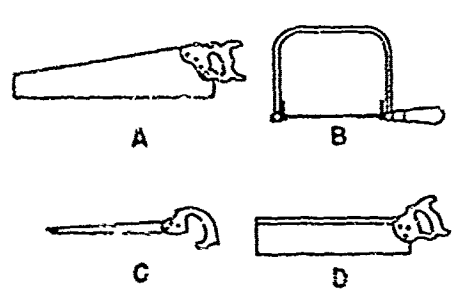
2. The wrench combination used to tighten the pipe connection is



3. The tool ordinarily used by a draftsman to transfer measurements from one drawing to another is



4. The tool ordinarily used in a miter box to cut angles is

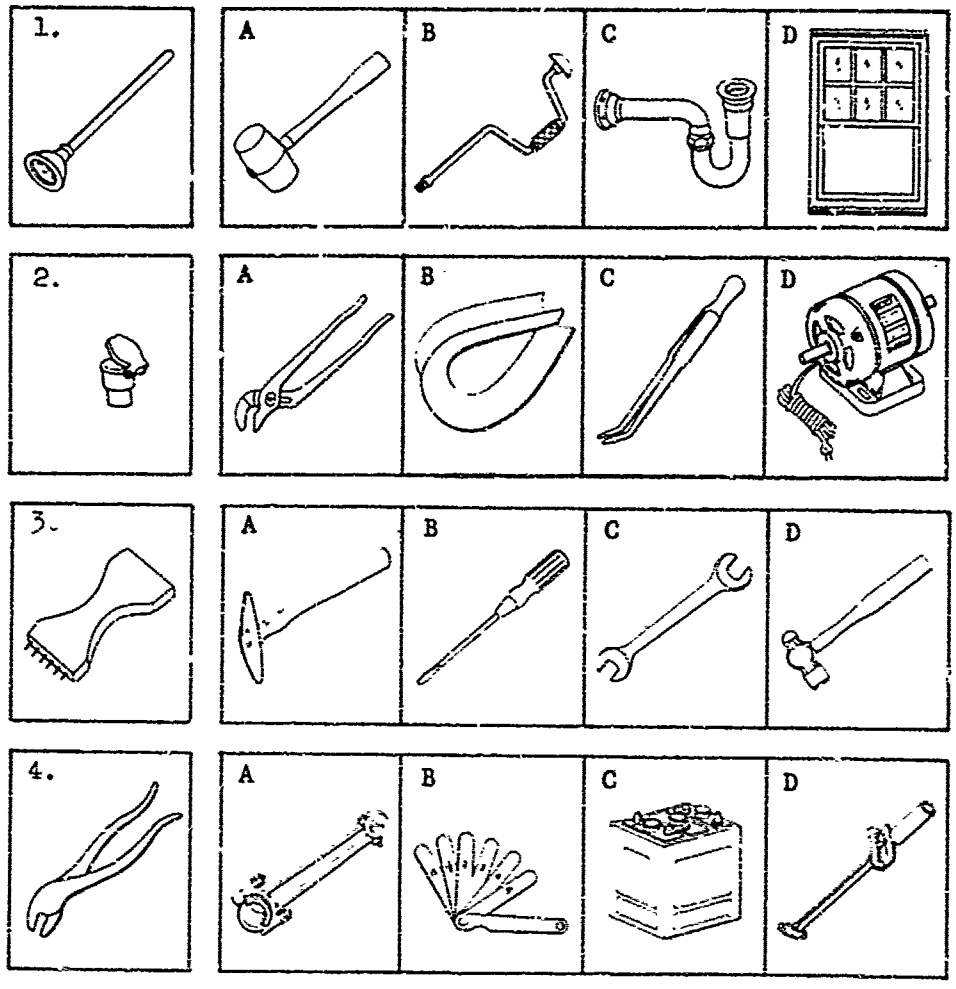


AFQT and AQB

SHOP MECHANICS TEST (SM): 25 items

SAMPLE QUESTIONS

Each item involves pictures of tools or other implements used in the shopwork of various trades. A question picture on the left is to be matched with one of four answer pictures to its right on the basis of the one with which it goes best.

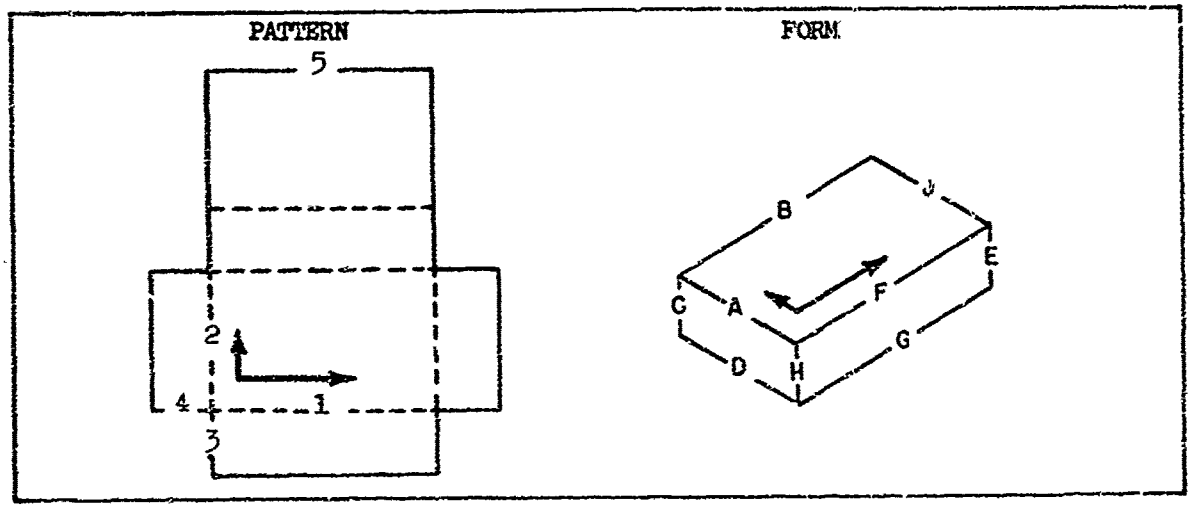


ACB

PATTERN ANALYSIS TEST (PA): 50 items

SAMPLE QUESTIONS

For a set of items, a PATTERN having numbered lines is shown next to a FORM having lettered edges. The examinee is to find which edge on the FORM would correspond to a particular line on the PATTERN if the PATTERN were folded where indicated. Arrows appearing in the same position on the PATTERN and the FORM serve as reference points.



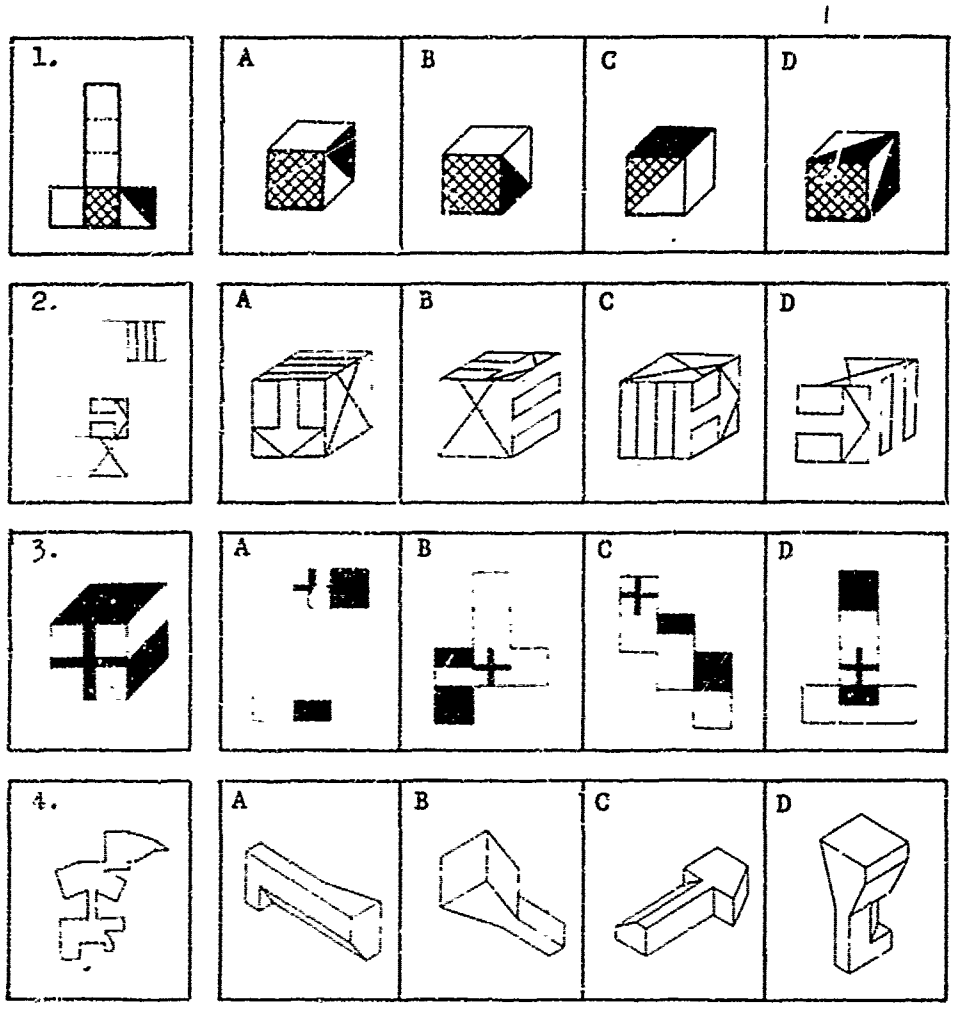
Pattern Line Number	Choose Corresponding Form Edge Letter
1	A,B,F,G, or J
2	A,C,D,H, or J
3	A,E,F,G, or H
4	C,D,E,G, or H
5	B,D,F,G, or J

AQE

PATTERN ANALYSIS TEST (PA): 25 Items

SAMPLE QUESTIONS

Each item is comprised of a question picture on the left and four answer pictures to the right. The question picture is either a two dimensional pattern that could be folded into one of the four figures shown to the right, or it is a three dimensional figure that could be formed by folding one of four patterns shown to its right. Some problems involve basic figures, such as cubes, with varying degrees of difficult surface designs and pattern layouts. Other problems involve increasingly difficult odd-shaped figures and patterns.



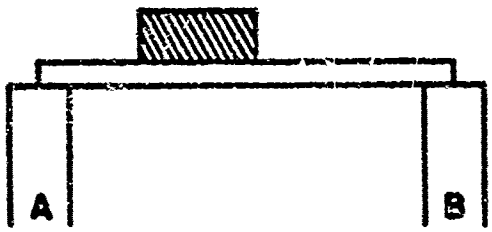
MECHANICAL APTITUDE TEST (MA)

SAMPLE QUESTIONS

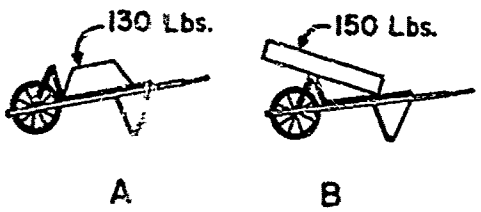
ACB: 45 items AQB: 20 items

Questions are asked about illustrations of mechanical problem situations that call for an understanding of physical principles.

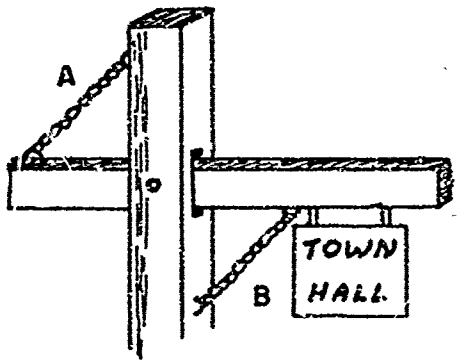
1. Which upright supports the greater part of the load?



2. Which wheelbarrow is HARDER to lift?

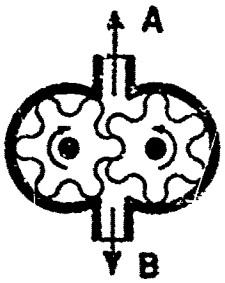


3. Which chain is helping to hold up the sign?



A) A
B) B
C) Both
D) Neither

4. The gears of this rotary oil pump fit very closely to the sides and end of the case and against each other. If the gears turn in the direction shown, which way is the oil forced?



AUTOMOTIVE INFORMATION TEST (AI)

SAMPLE QUESTIONS

ACB: 40 items AQB: 20 items

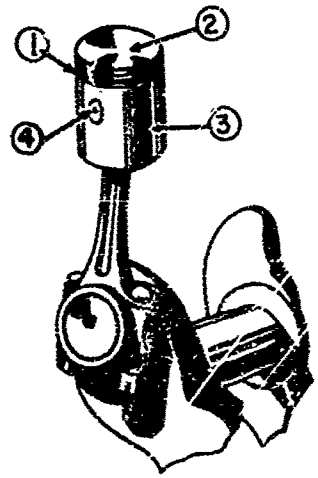
Items involve the identification and operation of automobile parts through descriptions and through pictures or diagrams.

1. Cross members are part of the

   A) shock absorbers
   B) frame
   C) hanger
   D) auxiliary spring

2. In the illustration below, which part is closest to the valves?

   A) 1
   B) 2
   C) 3
   D) 4



3. Jerky application of power to the crankshaft is eliminated by the

   A) crankcase
   B) main springs
   C) shock absorbers
   D) flywheel

4. The function of the rotor is to

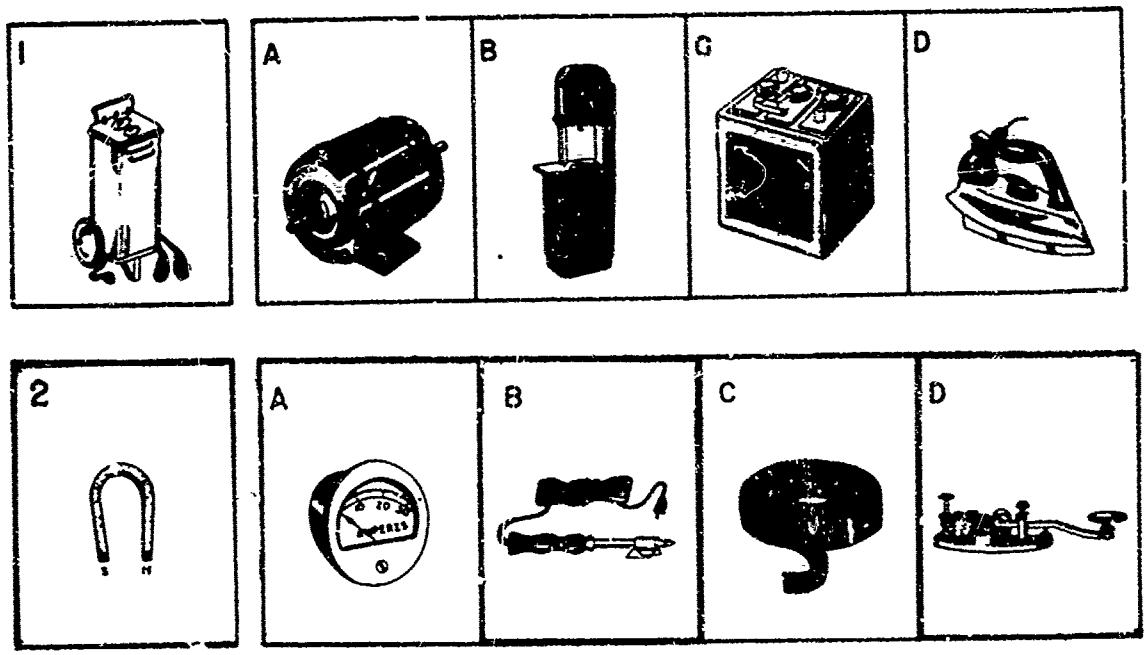
   A) open and close the distributor points
   B) rotate the distributor cam
   C) distribute electricity to the spark plugs
   D) rotate the distributor shaft

ACB

ELECTRONICS INFORMATION TEST (ELI): 40 items

SAMPLE QUESTIONS

This test contains an equal number of both verbal items and picture items calling for a knowledge of electronic principles. The picture items require the examinee to associate electronically a pictured object with one of four other pictured objects.



3. An electric switch is used

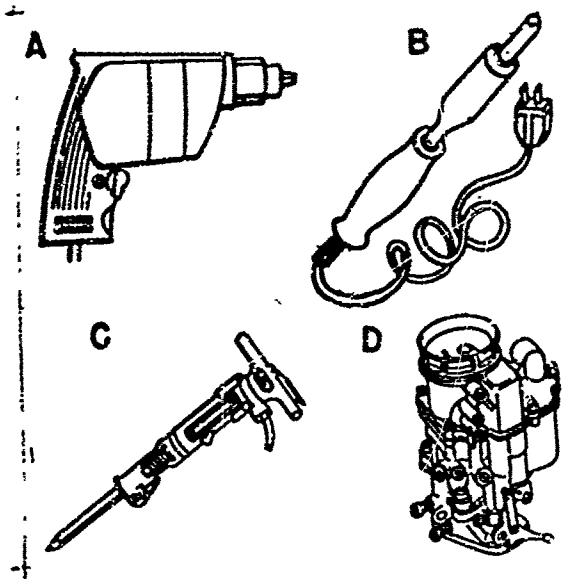
   A) only to break a circuit
   B) only to complete a circuit
   C) either to break or to complete a circuit
   D) to change the direction of flow in a circuit

4. The device used to conduct current from armature coils to the brushes of an a.c. generator is a

   A) catchall
   B) rotor
   C) stator
   D) slip ring

AQB
ELECTRONICS INFORMATION TEST (ELI): 20 items
SAMPLE QUESTIONS

1. Which contains an electric motor?

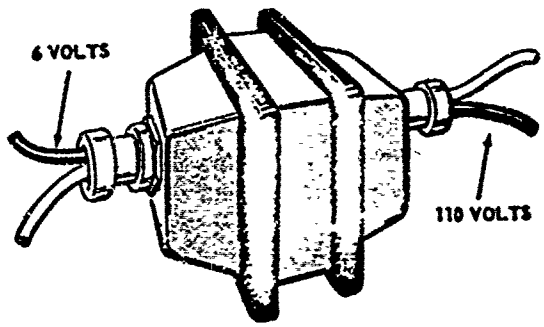


2. Mechanical energy is converted into electrical energy by a

A) generator
B) turbine
C) converter
D) motor

3. The figure below is

A) a circuit breaker
B) a transformer
C) an outlet box
D) a weatherproof receptacle



4. Usually, the moving part of a direct-current motor is th.

A) field
B) armature
C) brushes
D) frame

CLASSIFICATION INVENTORY (CI)

SAMPLE QUESTIONS

ACB: 125 items AQB: 75 items

The test consists of self-description items for which the examinee is to indicate the choice that most closely reflects his personal background, experiences, attitudes, and opinions of himself. The items were empirically selected to predict combat effectiveness.

1. The job of <u>explorer.</u>

   A) I think I would <u>like</u> the job.
   B) I think I would <u>dislike</u> the job.

2. Which would you like better?

   A) going to see a mystery movie
   B) going to a boxing match

3. I have never done anything dangerous for the thrill of it.

   A) Yes, this is true of me.
   B) No, this is not true of me.

4. Which describes you the LEAST?

   A) liberal with money
   B) look on all sides of a question
   C) religiously exact
   D) respectful
   E) smooth in movements

GENERAL INFORMATION TEST (GIT)

SAMPLE QUESTIONS

ACB: 50 items

AQB: 30 items

This is an objective test of knowledge in a variety of areas such as the Army, sports, guns, and automobiles.

1. The stripes on a soldier's sleeve that show how long he has been in the service are called

   A) crimp checks
   B) hash marks
   C) rockers
   D) hitch lines

2. The "No. 00 Buck" is correctly used in a

   A) .30 - 06 rifle against a moose at long range
   B) 16 - gauge shotgun against pheasants
   C) 20 - inch barrel carbine against a deer in dense cover
   D) 12 - gauge shotgun against a nearby bear

3. A regulation baseball diamond is a 90-foot square; a softball diamond is a

   A) 60 - foot square
   B) 75 - foot square
   C) 90 - foot square
   D) 120 - foot square

4. The intake and exhaust valve stems in an automobile engine are driven by the

   A) transmission
   B) crankshaft
   C) camshaft
   D) drive shaft

ACB

ARMY CLERICAL SPEED TEST (ACS)

ACB: 110 items AQB: 110 items

Number Reversal: 60 items

The examinee is to blacken the space under:

"R" (for RIGHT) if the numbers are exactly reversed,
or "W" (for WRONG) if the numbers are not exactly reversed.

1. 0874 R W 4730

2. 0159560 R W 0659510

Coding: 50 items

The examinee is to find the number that goes with the question word by looking for the word and number pair in the key, then mark his answer for each question. Question 3 is already marked.

KEY:

auto	2715	house	3451
bread	1413	train	2864

		1413	2715	2864	3451
3.	train			■	
4.	bread				
5.	auto				
6.	train				
7.	house				

ACB

ARMY RADIO CODE APTITUDE TEST (ARC)

The ARC is completely recorded on tape, including instructions, practice exercises, and examination questions. The first part of the test is composed of 270 learning exercises designed to teach the examinee the three letters I (two dots), N (a dash and a dot), and T (one dash). The exercises are presented at approximately 4 to 7 words per minute. Immediately after the learning exercises, a test of 150 items is given to measure how accurately the three code signals can be recognized at varying speeds. The first 75 items are presented at approximately 21 words per minute, and the second 75 at approximately 15 words per minute. Responses are marked on machine scorable answer sheets presenting three alternatives for each item.

APPENDIX C VALUE OF IMPROVED EFFECTIVENESS OF CLASSIFICATION TESTS

A new classification battery has been developed that, compared to the old battery, improves the accuracy with which potential for training performance is assessed. The old battery yielded a gain of about 55 percent of the maximum possible gain in performance that could be realized with perfectly accurate measurement. (Maximum possible gain is explained in the main body of the present report, pages 5 through 7.) The new classification battery yields a gain of 65 percent of the maximum, 10 percent higher than the old.

The increase can be translated to dollar value by determining to what extent training performance would be enhanced by using the new battery to make initial assignments to job training. This determination was accomplished through simulation studies conducted on a computer. Test scores on both the new and the old classification batteries were generated by the computer; the scores were representative of recent Army input. The men simulated were assigned to job areas in accordance with Army quotas for the different jobs. Each man was assigned by the computer in two ways--once on the basis of scores on the old battery and once on the basis of scores on the new battery. For each man, training performance was estimated in both assignments, and the expected performance summed for all men. A representative group of men assigned to jobs on a random basis (that is, with no knowledge of potential) are said to have an average expected performance score of 100. A representative group assigned on the basis of the old classification battery produced an average expected performance level of 102.5. When assigned on the basis of the new classification battery, the average performance level was 104.5. The gain of two points, from 102.5 to 104.5, can be translated to dollars by considering the cost and value of job training.

According to recent Army figures, the median cost of providing job training to produce a qualified worker in the field is $4,000. (While there is an additional cost of about $2,000 to procure a man and put him through basic combat training, this cost is not used in computing the value of job training performance.) Individual jobs vary substantially from the median of $4,000, but the figure is a reasonable average for the cost of job training. Since the average worker costs $4,000 to train, in that sense he is also worth $4,000 to the Army upon completion of job training. The below average man is expected to produce less than the cost of training, and the above average man is expected to produce more. Following this argument, anything that can be done to enhance the performance of the workers will increase the return to the Army of its investment in job training.

An unsatisfactory man is defined as one who yields no return to the Army from the investment in training; the $4,000 is lost. The person whose expected training performance is only 80, where the average is 100, is unsatisfactory. Using these definitions as a basis for calculation, only 13 percent of the population of young men liable for service would be unsatisfactory in the typical Army job; the 13 percent includes the 10 percent which by law is mentally unqualified for service. As the scale of expected performance is ascended, the return to the Army increases, until at the average level of 100, training cost and value of performance are balanced. Assuming a linear increase, each point increase between 80 and 100 is worth $200. ($4000 divided by the 20 points between 80 and 100.)

The scale also extends to the positive side. Each point of expected performance above the average is also worth $200. Since the new classification battery would increase the level of expected performance by two points, the value of the gain is $400 per man. With an annual input of 200,000 men, the Army would realize an additional $80,000,000 ($400 times 200,000 men) worth of increased performance by using the new battery to make training assignments.

The $80,000,000 is a gain that could be realized because of the increased effectiveness of the classification tests. This analysis does not imply that the increased performance would result in an immediate corresponding reduction in the Army budget. It does mean, however, that for a fixed number of enlisted men the overall quality of performance will be higher. As the Army force is reduced, each position becomes more important and the quality of each individual's performance more critical to the accomplishment of the Army mission.

APPENDIX D SELECTION AND CLASSIFICATION PROCEDURES

ENLISTED SELECTION AND CLASSIFICATION[1]

Selection and classification testing in the Armed Forces had its origin in World War I when the Army Alpha and Beta tests were used. During World War II, over 13,000,000 men were tested with the Army General Classification Test.

The Armed Forces Qualification Test (AFQT) was developed and introduced in 1950 as a screen for determining mental qualification for the armed services. Research has produced improved forms from time to time. During the 20-plus years that the AFQT has been used, the number of potential failures identified by the AFQT and kept out of the services is estimated at over 1,000,000. In recent years, the cost of accessioning an enlisted man and putting him through initial job training is estimated at about $6,000. Over 55,000 potential failures in the services were identified by the AFQT in calendar year 1970 and kept out of active duty. Thus, use of the AFQT to identify these potential failures in recent years has avoided annual accessioning and training costs estimated at $330 million.

Classification tests are given to enlisted men to determine the job areas in which the men are best qualified. The payoff from classification testing is improved performance because the men can be assigned to jobs where their aptitudes or potential to succeed are highest. Over the 20-plus years that classification tests have been used, an estimated 450,000 man-years of improved productivity has been obtained.

In recent years, the value of a man-year of productivity is about $5,000. In one year, the average estimated gain in productivity is 22,500 man-years; the annual dollar worth of classification testing is thus about $112 million.

The total annual gain in recent years from enlisted selection and classification is estimated at $442 million. Considering the magnitude of Department of Defense operations, the estimated gain of $442 million is probably an underestimate of the savings from selection and classification techniques. The alternatives to tests are random assignment or decisions based on education, work history, and subjective impressions regarding individuals. Tests are more likely to provide valid assessments of current potential to perform and are free of personal bias and subjective judgment.

1/ This statement was prepared by Dr J. W. Bowles, US Air Force, Dr. M. F. Wiskoff, US Navy, and Dr. M. H. Maier, US Army.

OFFICER SELECTION & RETENTION

The maintenance of a well-trained, effectively performing officer cadre is very expensive. For example, the cost to obtaining one commissioned officer through the service academies is $50,000. Once an officer is commissioned, there is considerable additional investment in professional training--an investment that is lost if the officer leaves the service. Personnel research over the past 20 years has greatly improved the capability to select and retain officer personnel.

Selection tests used on applicants for the service academies have identified an estimated 1200 potential failures per year. Screening tests are also used with applicants for Reserve Officers Training Course (ROTC) and Officers Candidate School (OCS). The annual savings in training costs realized through the use of these screening procedures is estimated at $50 million.

New techniques have been developed through research to enhance the overall performance of commissioned officers. Tests have been developed to identify ROTC applicants most likely to complete officer training and remain on active duty after completing their minimum requirements, and classification techniques are available to identify officers with greater potential in combat leadership or technical-managerial leadership. On this basis, officers can be assigned to the areas where their potential is high. The increased retention and utilization of commissioned officers resulting from use of selection tests and classification procedures are estimated to be worth $6.4 million per year.

PILOT SELECTION

Trends dating back to World War II have shown that, without the use of screening tests for aircraft pilots, rates of elimination from training average 50% or more of the input. Since the number of pilots required during World War II could not have been trained with such high elimination rates, the services undertook research to develop effective selection procedures for pilots. The effort was successful to the extent that elimination rates from primary pilot training during World War II were cut in half. Based on cost figures of the time, the World War II pilot selection program saved approximately $190 million in training costs.

Tests to select pilots for fixed wing aircraft and helicopters continue to be used by the services. Recent cost analyses (taking into account average time to elimination from training) show that each pilot training failure costs $16,000. The use of selection tests for pilots has resulted in an estimated savings of $32 million per year in training costs.