Property of: Hq USAF
AFPDPMC
Room 2315, T-8
Please return to: (Mrs Brunner
ox 69404)

AD681092

REPORT

of the

THIRD ANNUAL CONFERENCE

MILITARY TESTING ASSOCIATION

23 - 26 October 1961

Reproduced by the
CLEARINGHOUSE
for Federal Scientific & Technical
Information Springfield Va. 22151

CONTENTS

Title	Page
Foreword	1
Conference Officers	2
Agenda	3
Address to Conferees	4
Discussion Committee Reports *	
Multiple Uses of Job Knowledge Tests (Committee #1)	14
Validation of Job Proficiency Tests (Committee #2)	24
Development of Test Outlines (Committee #5)	28
Development of Test Items (Training of Item Writers) (Committee #6)	34
Automation in Processing Evaluation Instruments (Committee #8)	38
* (Former MTA Committees 3, 4 and 7 have been combined with the above or abolished)	
Activities Represented at the Conference	40
List of Conferees	43

BLANK PAGE

FOREWORD

During the period 27 - 29 October 1959, upon invitation from the Commanding Officer, U. S. Naval Examining Center, Great Lakes, Illinois, sixty selected military and civilian representatives from the Army, Navy, Air Force, Marine Corps, and Coast Guard met at the U. S. Naval Examining Center to discuss areas of common interest in the field of enlisted job proficiency evaluation.

This conference of military and civilian personnel who were directly involved in job proficiency evaluation provided such a fine opportunity for the interchange of ideas and for the discussion of new developments and techniques in this field that the conferees established the Military Testing Association (MTA), a non-profit activity, for the purpose of providing a continuing means for exchange of professional information and data.

It was resolved by the membership that the position of President, Military Testing Association, should be held alternately, on an annual basis, by Army, Navy, and Air Force representatives. It was also resolved that an annual conference of the membership and invited guests should be hosted by each of the Armed Services in turn.

The Second Annual Conference of the Association convened at Fort Benjamin Harrison, Indianapolis, Indiana, 20 - 22 September 1960, with the U. S. Army Enlisted Evaluation Center as host. This is a report of the Third Annual Conference, hosted by Personnel Laboratory, Lackland Air Force Base, Texas, 23 - 26 October 1961. The Fourth Annual Conference is being planned by U. S. Naval Examining Center, Great Lakes, Illinois, to be held there during September or October 1962.

CONFERENCE OFFICERS

Chairman

Colonel Fred E. Holdrege (USAF)

Arrangements Chairman

Lt Col Albert S. Knauf (USAF)

Program Chairman

Major Frank L. McLanathan (USAF)

Committee Coordinator

Mr. Jack W. Tinkle (USAF)

Discussion Committee Chairmen

Mr. Claude F. Bridges (USA)
Mr. Erling A. Dukerschein (USN)
Mr. John E. Partington (USA)
Mr. Raymond V. Baker (USAF)
Captain Ernest Madril (USAF)

AGENDA

24 October 1961

0830	Introduction	Colonel Fred E. Holdrege (USAF) Chief, Personnel Laboratory
	Welcome Remarks	Major General William J. Bell (USAF) Commander, Lackland Military Training Center
	Opening Remarks	Lt Colonel Charles C. Anderson (USAF) Chief, Evaluation Branch Directorate of Personnel Planning, USAF
	U. S. Army Enlisted Evaluation Center	Colonel Eugene E. Miller (USA) Commanding Officer, USAEEC
	U. S. Naval Examining Center	Captain H. H. Greer, Jr. (USN) Commanding Officer, USNEC
	U. S. Coast Guard Institute	Captain R. R. Curry (USCG) Commanding Officer, USCGI
	U. S. Air Force Personnel Laboratory	Lt Colonel Albert S. Knauf (USAF) Chief, Specialty Knowledge Test Branch
1115	Personnel Research	Dr. A. Carp (USAF) Technical Director, Personnel Laboratory
1345	Announcement of Committees	Mr. Jack W. Tinkle (USAF) Specialty Knowledge Test Branch
1400	Tour of Facilities	

25 October 1961

0830 - 1700 Committee Meetings

26 October 1961

0830	Committee Reports	Mr. Jack W. Tinkle (USAF) Committee Coordinator
1100	General Business Meeting	Major Frank L. McLanathan (USAF) Program Chairman
1130	Closing Remarks	Colonel Fred E. Holdrege (USAF) Chief, Personnel Laboratory

Address to Conferees

OCCUPATIONAL STRUCTURES RESEARCH
Dr. A. Carp, Technical Director, Personnel Laboratory

As the name implies, this research is concerned with Air Force jobs and job structures . . . . how to identify them, describe them, structure them, evaluate them, and determine their qualifications requirements.

This research enterprise was established at the direct request of Hq USAF and I must admit that the Laboratory was at first somewhat reluctant about accepting the assignment. Our hesitation was not because the importance of the mission was unrecognized, but because of the magnitude of the undertaking and our lack of insight as to how the necessary data could be collected, organized, and analyzed. In fact, there were real questions as to whether some of the problems were even amenable to "research" in the strictest sense of the word. Fortunately, things are not so black as they seemed; we feel we now have an ongoing research program with real potential pay-off value to the Air Force.

In order to save time, today I will limit my discussion primarily to the three problem areas which have the greatest relevance to this conference.

First I will discuss the development of methods for collecting, organizing, and reporting information describing Air Force jobs.

Second I will talk about the problem of work structuring.

Finally I will discuss the problem of on-the-job criterion development.

The first problem, that of developing methods for determining what personnel do on the job has obvious implications for many Air Force programs, one of which is the construction of specialty knowledge tests. To be of maximum utility to the using agencies, we decided that a procedure. for collecting job information must meet at least the following requisites:

1. Any such procedure must provide up-to-the-minute data in quantified form which is descriptive of the work performed by Air Force personnel.

2. The procedure must be economical to establish and maintain.

3. The method must be flexible so that it can be applied to small or large samples as often as deemed advisable.

4. It must be amenable to obtaining information from incumbents, supervisors or other knowledgeable individuals.

5. Data collection must proceed smoothly and mesh with primary work activities.

6. It must provide accurate information which reflects changes in job structures.

7. It is highly desirable that the resulting data be in a form which can be processed electronically with a minimum expenditure of manpower.

8. The data and the procedure itself must be amenable to changes dictated by operational requirements or research.

9. Last, but perhaps most important, it must be possible to determine dependable estimates of the validity and reliability of the obtained information.

In the search for a procedure that would most nearly meet these criteria, methods used in the major government agencies were surveyed and the literature in the area of occupational analysis was reviewed. A series of exploratory studies was conducted and the needs of Air Force using agencies were investigated. All of these sources of information led to the decision that some type of task inventory had the greatest probability of satisfying the requirements for a flexible, economical, reliable, and valid procedure for systematically collecting and organizing information describing Air Force jobs.

Since that decision was made, the Personnel Laboratory has conducted scores of studies concerning how task inventories should be constructed and administered, and how the resulting information should be organized and reported. A milestone was posted early this year when the Air Force officially adopted a task inventory job survey procedure which is outlined in Air Force Manual 35-2.

The procedure used by the Air Force for constructing inventories is straightforward. Using information derived from job descriptions, job

training standards, and other available sources, the initial duty and task list is assembled into a standardized format by following instructions which are spelled out in detail in the manual. At least two technical advisors assist in this initial construction in order to make the inventory as complete as possible and to insure the proper use of technical terminology. The tentative inventory is then sent out to five or more Air Force bases. At each base, one or more technical advisors, acting as subject-matter experts, carefully review the inventory. They modify or delete duty or task statements that are judged to be inappropriate. They also add significant duties and tasks which are known to be performed but which have not been included. Advisors are asked to pay particular attention to the use (or misuse) of technical terms and are asked to make suggestions or recommendations for improving the inventory coverage. Using this information, the final inventory is developed and administered to large samples in the field by local base test control officers. The incumbents are asked routinely to indicate the tasks which they accomplish as part of the job and how their work time is distributed across those tasks. In addition they may be asked on occasion for other information relative to these tasks, such as (1) frequency of performance, (2) amount of supervision required (or exercised), (3) time required to perform a task, (4) complexity of the task, (5) difficulty as compared with other tasks, (6) amount of experience required for effective performance and (7) importance of the task to the accomplishment of unit mission. The final selection of task attributes to be measured must be made in the light of the utility of the information to using agencies and the particular reasons for making a job survey.

Much of the research accomplished by Personnel Laboratory with respect to the task inventory has dealt with three problem areas:

1. The development of rating factors to use in conjunction with the task statements.

2. Evaluating the accuracy, reliability and completeness of information collected with inventories.

3. Developing efficient programs by which the information collected using inventories can be processed by the electronic computer and brought to bear on Air Force problems. I cannot discuss these in detail. Let me simply state that there has been nothing in our research findings thus far to discourage us from our approach. We have checked the accuracy of our results in many ways: (1) We know that if you ask an

incumbent to fill out an inventory on two occasions, he will give you essentially the same information twice, (2) supervisors agree to the reasonableness of information provided by their subordinates, (3) information obtained using daily work records agree in general with the information gathered using inventories, (4) when we list tasks known not to be accomplished they fail to get checked, (5) when we list the same task twice on the same inventory, slightly reworded, the incumbent usually provides the same rating values to the two stimuli, (6) low level incumbents rarely check high-level tasks -- and so on. We believe that we are getting high quality information using inventories.

Once we get task information onto magnetic tape we are in a position to perform a great variety of analyses. For example, we can provide information for revising training courses by locating those tasks which incumbents claim are giving them trouble; We can assist manpower officers in comparing two units by indicating how the work time of each is utilized; We can validate the accuracy of Qualitative Personnel Requirements Information; We can provide guidance for the development of new selection and classification tests -- I could list many more, but I will limit discussion to those uses which have an impact on our specialty knowledge tests. The first way the results of inventories influence these tests is through the official specialty descriptions. Changes in these descriptions may be made as a result of task inventory surveys, and these changes necessarily influence the specialty knowledge tests. Each question in these tests must be tied to some statement in the official descriptions.

However, it must be recognized that the official specialty descriptions are at best very general, and more detailed information is needed for the preparation of test outlines in order to make sure we have good content validity. For this reason, we are working on special computer outputs which are specially designed to assure proper content coverage of our specialty knowledge tests. For example, we will be in a position to list all of the tasks performed by incumbents in a specialty area, indicate the amount of specialty time spent on each of these tasks, and provide additional information when needed, such as the relative importance of the tasks to the accomplishment of Air Force mission.

GROUPING RESEARCH

Now let me change the topic for a moment and talk for awhile about the research we have been doing in the area of work structuring. In the

Air Force, work structuring takes place at three levels. First, tasks are grouped into jobs. Secondly, jobs tend to fall into homogeneous clusters, called job-types, which in turn are grouped into specialties. Finally, specialties are loosely assembled into broader occupation categories, called career or utilization fields. I will restrict my comments to the first two levels.

It isn't difficult to show the impact of job structuring on the efficiency of the Air Force. Consider, for a moment, effect of job engineering on the utilization of talent. The Air Force during the last few years has had a difficult time locating enough high-level talent to fill quotas for the rapidly expanding technical areas. Currently many such jobs require personnel in the upper 20% of the manpower pool. However, when one looks at these jobs, it becomes obvious that a significant proportion of incumbent time is devoted to tasks that any well-motivated airman with average talent could learn to accomplish. If the tasks which are assembled into jobs were made more homogeneous with respect to level and type of personnel requirements we could conserve use of incoming talent, classify more efficiently, and simplify training. Personnel Laboratory is working on this problem from several angles. For example, we are developing a set of "universal work requirement factors," which we hope will assist us in determining the similarity of tasks in terms of talent demands.

Perhaps the way that jobs are assembled into specialties is even more important, since the Air Force specialty is the basic unit that guides selection, classification, training, assignment and rotation of personnel. It is extremely important that the jobs included in a specialty be highly similar. To the extent they are not, the Air Force pays a big price in terms of money and efficiency. This becomes clear when one considers any one of a dozen Air Force policy areas, but for an example, let's consider the policy on airmen rotation. Airmen in the Air Force change jobs on an average of about once each 3 years, and can be moved freely from any one job to any other job having the same specialty number. In a sense then, management has built fences around similar jobs in order to control personnel reassignments. When an airman changes jobs, a major cost to the Air Force is the amount of time required for him to reach the same level of proficiency in his new job as he had attained in the job from which he was transferred. If the jobs included in a specialty are not reasonably homogeneous, the Air Force must pay in two ways: (1) At any given moment, there will be a large number of incumbents who have not reached proficiency in their job assignment, and (2) the cost of on-the-job training becomes excessive. It becomes clear, then, that jobs should be organized so th the cross-training time among jobs

within specialties is minimized; that is, so that the Air Force pays the lowest cost possible as airmen are rotated among the jobs in their specialty area. Fortunately, if specialties are structured in this manner, it can be shown that such action will also tend to simplify talent and training demands.

During the last year, the Personnel Laboratory has developed a computer technique which begins with a matrix of cross-training times among jobs and shows how the jobs should be organized into specialties so as to meet the above-stated criteria. At the moment we do not have a satisfactory method for estimating the cross-training time between jobs, but we believe we are making progress in this area.

An important advance made by the Laboratory during the last year has been the development of a method for identifying job-types using task inventory information. From one point of view there are as many jobs in a specialty as there are incumbents. However, we can never hope to deal with each job individually, while there are no restrictions on base commanders concerning the way they organize work in order to accomplish their mission, the fact is that two commanders having the same mission will often organize their work in somewhat the same manner. Thus the jobs on one base are often highly similar to jobs on other bases. We conceive of these similar jobs as being "job-types," and we feel it is important to identify and define them, for several reasons. First of all, we can more accurately define the work accomplished by incumbents in a specialty with job-type descriptions than we can with a single global description. A global description in reality does not define any single incumbent's job very well. A more important reason, however, is that job-types can be compared for homogeneity. If the work in a specific job-type is sufficiently different from the work in other job types within a specialty, then perhaps a new specialty or shred-out should be created. Before we can restructure Air Force specialties we must first determine the interrelationships among job-types. This has already been discussed. However, our first task is to identify and define job-types, and this looked like an impossible problem until recently. We now have a computer grouping program which appears to be very effective for identifying and describing job types.

For example, we might start with 250 job descriptions, each indicating how one incumbent's work time is distributed across the 300 tasks listed on a task inventory. Since each description covers 100% of a man's work time, we can account perfectly for the work time of the 250 individuals when we use the 250 separate descriptions. However, it is essential that

we reduce the complexity of this information by grouping those individuals together who have similar jobs. The computer accomplished this for us in a step-wise fashion. The first step involves locating the two men having the similar jobs and developing a single description which will account for their work time with the least error. In order to accomplish this, the computer actually develops 31,000 composite descriptions -- one for every pair of jobs -- and computes an error term in each instance. We accept merger of the two descriptions which results in the smallest error.

At the second step, the computer compares the possibility of combining some third description with the first accepted pair, and defining all three with a single description; or, as an alternative, it may combine two of the ungrouped descriptions. The choice is based on a comparison of the error associated with each of the 38,000 alternatives at this stage. This process is continued until we have defined the 250 jobs as accurately as possible with every possible number of description from 250 down to one. If, in order to reduce the number of descriptions during the latter stages of the clustering process we are forced to describe two somewhat different job-types with a single description, the error term will take a sharp rise. This helps us to determine the number of identifiable job-types in a specialty. At this stage we have the computer publish the best description for each job-type, and tell us the location and description of individuals working in these job-types. This is just one example of the way we utilize task inventory information in our study of work organization.

CRITERION DEVELOPMENT

Another problem being worked on by the Personnel Laboratory concerns the development of an acceptable on-the-job criterion composite. This is recognized as being one of the most important missions of the Laboratory. We currently select and classify personnel so as to maximize the chances of their succeeding in technical school training. We realize we should be selecting and classifying personnel so as to maximize their chances of succeeding on-the-job. We don't have even an acceptable intermediate criterion against which we can validate our specialty knowledge tests. In this case we simply take measures to assure good content coverage and acceptable internal test characteristics. Then we hope for the best.

The complexity of the job-criterion development problem makes it appear unlikely that we will have it solved in the immediate future. However, we

do feel that we are making progress, and for the first time are beginning to feel that a satisfactory solution can be obtained. I won't take time to elaborate on our efforts in this area, but I would like to state that our approach is rather unique. Our goal is to evaluate the effectiveness of incumbents working in a specialty area in terms of every characteristic which contributes to their over-all worth to the Air Force. If a man wastes the time of his associates, this may detract from his score. His job knowledge will contribute to his score only to the extent he applies this knowledge. In some specialties, the ability of an incumbent to deal effectively with other individuals may contribute heavily to his overall evaluation; in other specialty areas, this factor may receive no weight at all. We will not be concerned whether or not the factors which get weighted into personnel evaluators are predictable by current selection tests, training courses, or specialty knowledge tests. As a matter of fact, we are at present not even concerned as how we will obtain on-the-job evaluations for those characteristics which we identify as being relevant to individual effectiveness. Our first goal is simply to define such characteristics and to determine their relative weight in an overall composite score for each specialty.

A new procedure has been developed by the Laboratory during the past year which makes criterion research appear to be considerably more practical than it has been in the past. This procedure involves use of a "simulated sample," each member of which is defined in terms of all characteristics thought to be relevant to incumbent effectiveness in a job area. A board is then assembled and asked to study the characteristics ascribed to each individual in the simulated sample and rate his overall effectiveness. Multiple regression analysis techniques are then employed to determine the relative importance of each characteristic in terms of its contribution to the board's judgments of personnel effectiveness. It can be demonstrated mathematically that, if care is taken in assigning the characteristics, the accuracy of the regression equation will be as good as would be obtained from having the board rate real people in terms of measured characteristics. The beauty of the simulated sample approach is that (1) no money must be spent in collecting evaluation in the field, and (2) characteristics can be evaluated which we currently cannot measure accurately. The main pay-off for this type of study is that it tells us where we should concentrate our energy in the development of measures to be weighted into criterion evaluations.

I would like to point out that the information we collect using the task inventory has relevance to the criterion development problem. Normally,

a supervisor is asked to rate the proficiency of an incumbent in terms of how well he performs what he is asked to do. However, we have some evidence that supervisors assign the most capable airmen to accomplish the most difficult tasks. If this is the case, it is evident that any ultimate criterion designed to measure the worth of an airman should consider what he does as well as how well he does it. We can determine the difficulty of work performed by a particular incumbent using the information he provides on a task inventory.

Before leaving the criterion development problem, I would like to mention some of the work we are doing on rating behavior. It is inconceivable that we ever will be able to develop an ultimate criterion for most Air Force jobs which does not include scores obtained from judges of one sort or another. This being the case, we have been attempting to systematically identify the types of errors and biases which affect judgment values, and are developing ways of eliminating or controlling these factors. I won't attempt to discuss these individually, but you might be interested to know that we can identify some raters who consistently give low ratings, and others who consistently give high ratings. We have also discovered that certain factors identified by psychologists working with human judgments of weight, light intensity and size also affect a supervisor's judgment of his subordinates. For example, an incumbent with average talent will be rated lower by his supervisor if he is working in a group of highly capable incumbents than he will if he is working in a group composed of less capable individuals. This phenomenon is known as the context effect. We must identify and learn to correct for such errors before we can hope to develop a reliable on-the-job criterion.

JOB EVALUATION

One last area, which I will discuss very briefly because it is not germane to this conference - The concern of proficiency testing is with the measurement of people. On the other side of the coin, we need to address ourselves to the problem of measuring specialties. It is not enough to determine that an airman possesses a certain degree of proficiency, we must also assure that his proficiency level is commensurate with the requirements of the job to which he is assigned. Assessing job requirements is the province of job evaluation.

Personnel Laboratory has devoted some of its research effort to the study of the judgmental processes involved in the estimation of job requirements. Work has also been done on the development of methodology which most economically and efficiently enables the collection of job

evaluation judgments. This work has led to a wide variety of activities including the development of new statistical methods, scaling techniques, semantics studies, industrial job evaluation practices, the economics of differential compensation, and other problems. Some of the implications of more extensive applications of job evaluation in the Air Force are intriguing and are currently one of our primary interests in this area. A most intriguing problem facing us now is the relationship between grade and pay. We suspect that, because of the need to retain skills, we have "over graded" certain skills at some long range cost.

Again, thank you for your attention, and I hope this account of Occupational Structures Research will be of some value to the conference.

Report of Discussion Committee #1

Multiple Uses of Job Proficiency Tests

Participants	Affiliation
Mr. C. F. Bridges, Chairman	USAEEC
Lt Cdr S. B. Russell	USCGI
Mr. C. J. Macaluso	USNEC
Mr. E. M. Ramras	BUPERS, USN
Capt S. Henley	OTSG, DA
Mr. N. J. Allred	USASA
Mr. B. J. Foley	USASATC&S
Dr. J. D. Teller	Hq USAF
Maj J. R. Shafer	Hq AFSC
Maj J. J. Felthaus, Jr.	Pers Lab, USAF
Dr. F. Harding	Pers Lab, USAF
Mr. H. L. Madden	Pers Lab, USAF

Problems Discussed:

Because of its crucial importance, and the urgency of extensive planning and action prior to any actual need, the mobilization and related uses of achievement tests were given most attention during the deliberations of the committee.

The appropriateness of the term "job proficiency tests" in the title of the committee was discussed briefly. The Air Force recently changed the title of their tests from "proficiency tests" to "knowledge tests." The pertinent dictionary definitions are similar and equally appropriate. However, the term proficiency both in its technical definition and in its popular usage is more inclusive, referring to the information, manual skills and mental skills making for expertness in an occupation. In defining "knowledge" Webster gives the term "practical skill" as synonymous with one of the usages of the term. However, the tenor both of the technical definitions and of popular usage, both tends to emphasize informational aspects in contradistinction to manual skills involved in the job. It was pointed out that the US Army Enlisted Evaluation Center does include evaluations of manual skills when evaluating proficiency in those jobs in which they are most critical determiners of proficiency, and

currently the Center is exploring techniques which may make practical such evaluations in most jobs for which they are warranted. In addition, for a majority of military specialties possession of information and mental skills are predominant among the determiners of proficiency. Hence, the more inclusive title is most appropriate for the Army tests. Since pencil-and-paper tests are used primarily as the basic instrument for the evaluation of achievement by the agencies represented, on the committee, and also since they are the ones proposed for the mobilization use to which the committee directed most of its attention, this committee concluded that the term "knowledge tests" was most appropriate for this report.

The deliberations, conclusions, and recommendation with respect to each of the eight problems discussed by the previous committees were reviewed and completely indorsed as being technically sound and very desirable. In discussing the diagnostic uses (Problem 8) the additional point was made that a minimum of 20 - 25 test items for an area are required for an acceptable degree of precision and validity. In addition, it was pointed out that much greater reliance could be placed upon differences between area averages for groups of individuals. The concensus of opinion of the committee was in agreement with these points. The conclusions and the general recommendation are presented as Annex 1 for ready availability to those considering this report.

Deliberations were then devoted to developing a detailed procedure for carrying out the general recommendation of the previous committee, which pertained to classification from civilian manpower pool directly into military specialties, both to obtain highly trained personnel for critical military specialties in time of peace and to effect optimal rapid utilization of shortage abilities in time of either partial or full mobilization.

The experience of the military service during World War I suggests and the current experience of both the Air Forces and the Navy, clearly demonstrates, that the answer to the question "Can military job proficiency tests be used effectively for classification from the civilian manpower pool directly into military specialties for selective recruitments, partial mobilization, and full mobilization into both the military services and industry?" should be an unqualified "Yes" with some modifications of current related procedures and materials.

In spite of their limitations, the United States Employment Service "Oral Trade Tests" were useful in WWI for the verification of experience and

training claimed by inductees. However, these short oral tests were out of date by WW II. Their low reliability, their lack of applicability to the most critical military specialties, and the great strides made since in measurement techniques all suggest that better instruments could now be developed which would be most useful for verification of claimed experience and training and for reliably identifying individuals possessing at least an acceptable minimum of critically needed knowledges and skills.

The soundness of this concept has been demonstrated by the experience of the Air Force with the results of their by-passed specialists. Their procedure permits immediate assignment to an Air Force specialty without the Air Force school training. The Navy Training Command at Memphis currently is reported to be using the Naval Examining Center's promotion examinations to identify candidates who can successfully skip to higher levels of training. In addition, at the last Military Testing Association Conference, Commander Callahan reported that the Navy has been very successful in selective recruiting for special higher skills - - making provision for the men to receive their required military training on-the-job. They use the man's civilian training as the primary basis for such selective recruiting, but use of tests would increase the number of potential selective recruits and would provide better assurance of using such special classification procedures only for qualified individuals.

The practical feasibility of economically and rapidly developing special tests both to aid in meeting current shortages in critical military specialties and in mobilization as a basis for optimal utilization of critical manpower both in civilian and military activities is enhanced by the fact that the test development agencies of the various services already have items, that can be utilized for this purpose. Each service has for each specialty area substantial banks of appropriate general, technical, and theory items with demonstrated effectiveness.

Attention was given to the fact that the possibilities inherent in data processing equipment can simplify the test development process and the analyses of the data in refining the test and establishing minimum cut scores as well as expand the procedures possible for precisely identifying the specific tasks and military specialties basically similar in two or more of the services. It was reported that the Military Occupational Classification Project of the Office of the Secretary of Defense in 1949 - 1951 developed statistical techniques for determining groups of military specialties common to the services and for quantifying their criticality. In fact, the noteworthy technique of analyzing tasks which was reported to the MTA

by Dr. Carp appears to be a development from the application of some of the iterative techniques devised for this OSD project. Perhaps these statistical techniques can be applied for this limited but crucial project. In addition, it would be possible to determine the relationships between all the military specialties and the larger classification of civilian jobs such as the USES Dictionary of Occupational Titles. Preliminary work along this line has already been done by the US Army Enlisted Evaluation Center for the Army MOS. This could serve to define similarities between the military services as well as for many other desirable uses. Other acceptably sound procedures also are feasible.

The tests thus developed would be used to screen-out special recruits and inductees that require a minimum of training to function effectively at higher skill levels, i.e., to screen-out men who are needed and can be used almost immediately. In many instances they would be by-passed and go directly to on-the-job training.

The tests might be administered and used by recruiting agencies until mobilization. Appropriate procedures would make them helpful in recruiting experienced men in critical shortage specialties.

After mobilization the tests might be administered and used at induction type centers.

In closing the deliberations, the point was emphasized that:

1. This project is envisioned as a very limited area of cooperation between the services for this critical purpose.

2. In the event that an all-out nuclear attack destroys our first strike capabilities, our second strike capabilities will have destroyed the enemies' power. In regrouping our forces, effective mobilization of civilian specialists becomes of crucial importance.

General Summary of Conclusions:

A. Primary Conclusions

(1) The basic tasks of many military specialties (NECs, Navy Enlisted Classification; AFSs, Air Force Specialty; and MOSs, Military Occupation Specialty) are very similar.

(2) It has been demonstrated by two of the military services that evaluation procedures can identify individuals who can function satisfactorily at a skill level higher than the normal entry level of personnel coming from the civilian population.

(3) It appears to be practical for the test development activities of the military services to develop common tests which will function effectively for all the services in identifying individuals in civilian life for direct assignment to a military specialty at an advanced skill.

B. Secondary Conclusions

(1) The basic tasks performed by reservists are either identical or very similar to the tasks performed by personnel on active duty.

(2) Two of the services have demonstrated that the same methods of evaluation will function effectively for both reserve and regular personnel.

(3) It appears to be desirable that all reservists not on active duty should be evaluated periodically by the same instruments as the regular forces in order to insure readiness for mobilization. (This is now accomplished by the Air Force for its ready reservists.)

General Recommendations:

A. Due to the need for rapidly integrating civilian job trained and experienced personnel into armed service jobs during mobilization, it is recommended that all of the services at the earliest possible date institute cooperative programs designed to develop tests for this purpose.

B. It is recommended that the military services explore cooperatively the utility in maintaining readiness for mobilization by periodically evaluating reservists with evaluation instruments developed for the regular forces.

Specific Recommendations: In order to implement the preceding general recommendation A, it is recommended that the following specific actions be taken:

A. Identify the common specialties among the services and civilian jobs and training which might prepare for these groups.

B. Determine the relative criticality of the specialties that are included in each of the common groups.

C. Conduct a study on several of the more critical groups of specialties in order to definitely determine the feasibility of developing common tests for use by all the services in identifying civilians for direct assignment to military specialties at advanced skill levels.

D. If the development of common tests is found to be feasible, assign specific responsibilities for each group to a test development activity. In making the assignments, special capabilities of each service should be exploited to the maximum extent possible.

E. The test development activity responsible for each test will utilize representative test psychologists and/or subject matter specialists to insure adequacy of the common tests for all the services.

F. After completion of each draft test, it will be coordinated and reviewed by specialists in each of the services and any necessary revisions made.

G. If mobilization does not seem to be immediately imminent, the initial test will be printed and distributed to the test development agency for each of the services to be administered to their servicemen for the establishment of minimal cut scores for their appropriate military specialties and skill levels.

H. If mobilization is judged to be imminent before more refined procedures are possible, item statistics will be used to determine cutting scores as precisely as possible. These cutting scores will be revised when sufficient data becomes available.

I. The developing agency will prepare manuals and other accessories along with the master copies of all materials for printing and distribution to all recruiting activities. These tests will be utilized by recruiting activities as authorized by appropriate regulations.

J. In the event of mobilization, these tests will be printed and distributed to induction centers for use as authorized.

ANNEX 1, REPORT OF COMMITTEE #1

CONCLUSIONS AND GENERAL RECOMMENDATION OF PREVIOUS COMMITTEE #1 WHICH WERE ENDORSED BY CURRENT COMMITTEE

1. Conclusions re Problem 1

Problem 1: How should job proficiency test scores be used as a basis for awarding proficiency pay?

Conclusions: All possible information must be given to enlisted personnel about the objectives and detailed workings of the system used by the service.

In well established military specialties, proficiency pay normally should be awarded to personnel above average in proficiency in their military specialty.

Very high successive levels of proficiency (such as those indicated by very high cut-off scores on the proficiency test) should be required for the higher P2 and P3 awards.

2. Conclusions re Problem 2

Problem 2: How can objective minimal standards for qualification in each military specialty be established?

Conclusions: The minimum fundamentals approach to setting minimum qualifying scores is impractical and technically unsound.

Minimum qualifying scores for a military specialty should be set by obtaining appropriate data and applying statistical research techniques.

A minority conclusion was that successful school graduates of MOS awarding courses should automatically be considered qualified for approximately a year. Enlisted personnel awarded a military specialty on the basis of on-the-job training should receive a provisional MOS subject to confirmation of qualification upon successful attainment of at least the minimum qualification score on the first MOS test administration immediately following award of the provisional MOS.

3. Conclusions re Problem 3

Problem 3: How can job proficiency tests contribute to more effective on-the-job and school training of enlisted personnel?

Conclusions: Job proficiency tests can be used to identify individuals in need of further broad training either on-the-job or in a service school.

Comprehensive job proficiency evaluations should be more useful than aptitude test scores in predicting future proficiency at a higher level.

Job proficiency evaluations can be used appropriately to evaluate the results of training given on-the-job, and in some military specialties, if handled correctly research-wise, of training given by the schools.

If content areas of a test are to be used to identify areas of strengths and weaknesses, a substantial number of items must be included for each area and the statistical significance of the differences must be provided.

4. Conclusions re Problem 4

Problem 4: How should scores be used to identify enlisted personnel most likely to succeed if promoted to a higher grade?

Conclusions: Comprehensive, valid measures of job proficiency at a given skill level are better predictors of future proficiency at the next higher skill level than specific measures of either aptitudes or a separate component of the job.

Competition for promotion should be service-wide.

5. Conclusions re Problem 5

Problem 5: Can job proficiency tests be used to establish minimum standards for reenlistment?

Conclusions: Job proficiency evaluations should be used, rather than aptitude test scores, to set minimal standards for reenlistment in the same or a closely similar military specialty.

The score above the absolute minimum which would be considered the current effective minimum would depend upon factors peculiar to the military specialty, such as immediate and anticipated needs of the service, amount of training required, and availability in the civilian manpower pool.

6. Conclusions re Problem 6

Problem 6: Can military job proficiency tests be used to establish service-wide selection and placement, and/or promotion standards for civilian jobs that parallel military jobs?

Conclusions: The overlap is so great that the regular proficiency test developed for military jobs can be used for parallel civilian jobs with little, if any, loss in effectiveness.

The services can, to their mutual advantage, collaborate in the development or exchange of tests suitable for civilian employees in selected jobs.

7. Conclusions re Problem 7

Problem 7: Can military job proficiency tests be used effectively for classification from the civilian manpower pool directly into military specialties for selective recruitment, partial mobilization, and full mobilization into both the military services and industry?

Conclusions: The soundness and effectiveness of using classification from the civilian manpower pool directly into selected military specialties have been demonstrated by the experience of both the Navy and the Air Force in their respective programs for selective recruitment and by-passed specialists.

Appropriate job proficiency tests would be the most effective practical method of accomplishing such immediate classification.

Both as a means of attracting critically needed specialists and to provide a sounder basis from which to plan mobilization under various conditions, arrangements to expand the use of such selective by-passing of recruits into specialist classifications are highly desirable for selected military specialists.

In the event of mobilization, the test development centers of the military services can play the role they should, only if appropriate prior planning is accomplished.

Job proficiency tests can meet a definite need in full mobilization. Careful study is necessary to determine how, and the extent to which, they will be used under varying conditions of mobilization.

The personnel requirements for essential civilian production, for essential civil services, and for each of the military services must be considered in

making plans for the use of proficiency tests during periods of partial or full mobilization.

8. Conclusions re Problem 8

Problem 8: Can job proficiency tests be used as diagnostic instruments?

Conclusions: Use of existing tests, utilized by all the services for both proficiency pay and advancement, for diagnostic purposes or profile scores is feasible if the following limiting factors are considered:

1. For predictive purposes the validity and reliability of part scores must be empirically ascertained.

2. For normative purposes reliability of the part scores is essential.

3. For nonpredictive and nonnormative purposes, such as individual motivation and individual remedial action, neither reliability nor validity is critical.

General Recommendation:

In view of the potential importance of multiple uses of job proficiency tests to the national defense efforts, it is recommended that this committee be continued and that an appropriate individual from the proficiency test development activity of each service be designated to continue cooperative exploration of the problems pertaining to classification from the civilian manpower pool directly into military specialties and other uses of proficiency tests during periods of partial and of full mobilization.

Report of Discussion Committee #2

Validation of Job Proficiency Tests

Participants	Affiliation
Mr. E. A. Dukerschein, Chairman	USNEC
Col H. L. Parris	Behavioral Sci. Lab, USAF
Capt G. E. Cowan	Pers Lab, USAF
Capt G. K. Cantrell	Pers Lab, USAF
Capt D. H. Taylor	Pers Lab, USAF
Dr. A. Carp	Pers Lab, USAF
Dr. J. E. Morsh	Pers Lab, USAF
Dr. C. J. Judy	Pers Lab, USAF
Mr. B. M. Vitola	Pers Lab, USAF
Dr. J. D. Teller	Hq USAF
Lt K. D. Albritton	USCGI
Mr. C. W. Hodges	BUPERS
Mrs. G. K. Shutter	USNEC
Mr. E. M. Sait	USAEEC
Dr. J. Boyd	USA Ord GM Sch
Mr. C. L. Legere	USA Security Agency
Mr. N. Yannuzzi	USA Ord

Problems Discussed:

The committee discussed the following six problems related to validation of job proficiency tests:

Problem 1: Why validate job proficiency tests?

Discussion: There was general committee agreement that validation of tests is necessary. The discussion was centered on "content validity." Members of the Personnel Laboratory reviewed some of the work they have been doing in the "job analysis area." They have designed a method that combines the open-ended feature of the questionnaire, the structure and organization of the check list, and the provision of verification of the observation interview. Initially, a duty and task inventory is constructed from references such as job descriptions, job training standards and in consultation with subject matter experts.

The preliminary inventory is then reviewed by technical advisors at five or more Air Force Bases. The revised inventory is then assembled in a booklet form and administered to large samples of job-incumbents who respond to task statements organized under major functions of their jobs, and add statements not listed. The incumbent's background information and data descriptive of tasks performed are also obtained. The selection of particular task attributes to be measured depends upon the requirements of the using agencies. A more complete explanation may be found in a paper entitled, "A Method of Job Analysis for the Air Force."

The committee concluded that "content validity" is important in assessment and that such validity for test construction was initially determined during the "job analysis."

Problem 2: The committee next discussed the problem of assessment and predictability.

It was generally agreed that assessment was the major purpose for which achievement tests were used, and that the "content validity" of a test is important in this area.

Some members of the committee felt that prediction might be a side product of assessment tests and that it might be well to attempt to measure this characteristic even if it is not done as a routine part of test analysis.

In this connection the American Psychology Association's position advocating predictive validity in tests was considered. In terms of service procedures test results may be sent to field units in some form of ranking rather than on a pass or fail basis. It was felt that in some cases this would lead to the use of these test results as predictors. In other instances, the use of test results as a basis for qualification introduces a secondary type of prediction if unqualified personnel are not considered for advancement. The possible advantage of developing tests with predictive validity was discussed. Some of these advantages were "objectivity," a "uniform" rather than a variable measure and construction of paper-and-pencil tests with known correlations with one or more performance tests.

The committee concluded that assessment was of primary importance in service tests and that more attention might be given to their "predictive validity" aspect.

Problem 3: The third problem discussed concerned what kinds of criteria may be used in validating proficiency tests.

One of the criteria considered was the rating scale. In the area there was some dissatisfaction with the results when rating scales were used. In some cases chance correlations were obtained.

Some possible reasons for these results were considered. It was noted that without variance in ratings, they can be of little use in validation. One suggestion concerned the problem of how completely the rater observed the performance of the individual being rated. Another possible reason concerned the loyality of supervisors for their subordinates as well as their responsibility for training them, and consequently, their responsibility for the proficiency of their subordinates. Another reason was the difficulty supervisors might experience in separating job knowledge from job performance.

In connection with attempts to improve the usefulness ratings a study was reported in which rating scales were correlated with end-of-course performance tests. Three types of raters were used; a peer group, the immediate supervisors and higher supervisors. Five rating scales were used which varied from the specific (job knowledge scale) to the general (overall value to the service scale). Correlation coefficients varied from .24 to .35 with a tendency for peer groups and the more specific scales to show the highest correlations.

The committee also noted the possibility of job performance tests and mock-ups for testing trouble shooting procedures as criteria for validation. Although these methods were felt to be substantially more expensive than ratings, their use to validate even a few tests would result in increased confidence in other tests constructed under similar methods.

The committee concluded that in the use of rating scales more attention should be given to the amount of time the supervisor has spent in observing the person being rated and the number of tasks that have been observed. It was further noted that reliability in ratings was desirable and might be improved by the use of multiple ratings. The committee felt that the use of rating scales geared to specific areas might be of value.

Problem 4: Who should accomplish ratings?

The committee considered three possibilities. Supervisor ratings, peer group ratings, and subordinate ratings. The committee briefly discussed

methods of locating "good" raters in terms of the amount of agreement between their ratings and various outside criteria and their use, as a possible line of investigation.

No firm conclusions were developed concerning this problem.

Problem 5: What role might factor analysis play in the validation of proficiency tests?

The committee briefly considered the value of factor analysis in the validation of proficiency tests. It was suggested that such analysis might be helpful in determining grouping of items into categories such as rote memory or reasoning type items as well as its use in modifying the job analysis. In this connection apparent weighting versus actual weighting of knowledge areas in a test was discussed.

No firm conclusions were reached on this problem.

Problem 6: Of what value would item validation be in the validation of proficiency tests?

The committee discussed the problem of item validity. It was felt that this type of validation could be determined by internal measures such as an item discrimination index, and might be accomplished through the use of a proportion of old items in each test with known item characteristics - in a sense, the use of pre-tested items.

The possibility of developing a completely controlled examination through the use of pre-tested items only, with a known relationship to a set of standards was briefly mentioned.

The committee concluded that item validation would be useful in validating proficiency tests and that internal measures formed a method of item validation.

The validation of proficiency tests continues to be a large and complex field of investigation. General agreement within the committee appeared in some instances and clarification of issues in others. The committee recommended continued study of this area.

Report of Discussion Committee #5

Development of Test Outlines

Participants	Affiliation
Mr. J. E. Partington, Chairman	USAEEC
CHSPCK R. V. Bercaw	USCGI
Cdr M. R. Rogan	USNEC
Mr. L. W. Johnston	USNEC
Mr. I. J. Newman	Pers Lab, USAF
Mr. L. D. Vannest	Pers Lab, USAF
Mr. C. B. Haradon	Pers Lab, USAF
Mr. G. S. Long	Pers Lab, USAF
Mr. C. Lutz	Pers Lab, USAF
Mr. C. J. Pearce	Pers Lab, USAF
Mr. A. P. Barra	LMTC, USAF
TSgt T. G. Herring	LMTC, USAF
Mr. H. M. Hyman	LMTC, USAF
Mr. J. D. Jones	LMTC, USAF
Mr. J. A. Naranjo, Jr.	LMTC, USAF
Lt R. W. Kielman	USAMFSS
Maj E. Louis	TAGSUSA
Capt J. F. McGraw, Jr.	USA Inf Sch
Capt A. A. Mesten	USA AD Sch
Mr. L. Rose	USA Trans Sch
Capt R. O. Rowland	USA PMG Sch
Mr. A. W. Weisbrod	USA Eng Sch
Ens B. K. York	USNEC
CWO R. H. Granger	USAOGMS

Problem 1: "Characteristics of a Test Outline and Factors Affecting Test Outline Development."

Discussion: The test outline is of paramount importance in test development. It is the blueprint or plan for the test. A good outline cannot assure a good test but a poor outline makes it very difficult to develop a good test. A test designed to test the knowledge of a job and the application of that knowledge to the job must be based on an outline which reflects accurately the various requirements of the job and the respective weights placed on each job

requirement. In reflecting these job requirements it is necessary that they be broken out in sufficient detail that the variety of accomplishments or expected outcomes will be included and none will be omitted. In accomplishing this, it is necessary that all possible sources of information be explored and that both subject-matter specialists and test development specialists participate. Test outlines must be based on official job descriptions and other official documents. These official source materials, however, must continue to keep pace with the job as performed. The job as performed must conform to that which is conceived by personnel management as the most desirable from the standpoint of the overall functioning of the organization of which the job is a part.

The Navy in general has a detailed breakout for each of its occupations in the "Manual of Qualifications for Advancement in Rating." This does not mean however that it should always be used without further breakouts and clarification when test outlines are developed.

The Air Force has available "Job Training Standards" which are a detailed breakout of job requirements. These standards, along with specialty descriptions and package training programs, form a basis from which the test outlines are developed.

The Army uses job descriptions, job analysis, and numerous manuals pertinent to each job in developing test outlines. There is no detailed breakout of job requirements available to the test psychologists at the Enlisted Evaluation Center for test outline development as the other services have.

Conclusions:

1. Each test outline should consist of a detailed breakout and weighting of the duties, skills, and knowledges required for successful performance of the job for which a test is to be developed.

2. Each test outline should assure that the job is fully covered but it must not include material which is not required according to official sources.

3. Each test outline should be "job oriented", not "academically oriented."

4. A test outline should reflect a "two-way" breakout. That is, expected goals and areas in which these goals are to be carried out should be reflected in each test outline.

5. Test outline development is a cooperative process. Subject-matter specialists, personnel acquainted by experience with the job, and test development specialists can share in the development and revision of test outlines.

Recommendation:

Job information be kept current and changes be made available for use in test outline development at the earliest possible time. Each service should make available to the test development agencies a detailed breakout of job requirements for each specialty. This would make available a standard set of requirements for use in test development; training-formal, on-the-job, and personnel classification and assignment.

Problem 2: "Item Types of Types of Ability Measured and Their Relation to Test Outlines."

Discussion: The type of test to be developed will determine to some extent the types of items to be developed. This in turn has some bearing on test outline development. Each service develops tests which are designed to test knowledge and also the application of that knowledge to the job. There are at least five types of items which can be considered when developing proficiency tests and these require differing abilities on the part of the examinees. (1) A vocabulary item asks for meaning; (2) A fact item asks about specific information - who, when, what, where; (3) A generalization item asks about a law or method of procedure; (4) An understanding item begins with "Why" or asks for "cause;" (5) An application item presents a problem. The language of the test outline can assist in guiding the item writer in developing the various types of items required for evaluating personnel.

Conclusions:

1. Personnel at lower levels of skills should probably be tested more with factual and generalization type items.

2. Personnel at higher levels of skill should be tested more with understanding and application type items.

Recommendation:

The language of each test outline should be such as to guide item writers or suggest to item writers the types of items to be written. Outlines should be

"action" oriented. Statements should begin with verbs - assemble, disassemble, repair, install, interpret, estimate, construct, applies knowledge to, understands, etc.

Problem 3: "Areas of the Job as Described in the Test Outline which can be measured by Performance Tests or by Performance Check Lists."

Discussion:

There may be one or more areas of a job which can be measured by performance tests or performance check lists. Such an area may be such that a paper and pencil test is not particularly suited for the purpose. Evaluating how an individual actually performs on the job may be easier to determine if a performance check list could be used for obtaining more objective information about the performance. For example, an Army Commander and his supervisor must evaluate the performance of Army personnel under their command with a Commander's Evaluation Report. Certain portions of this report would probably be more accurate if a performance check list were available for "checking out" the performance of individuals being rated.

The Navy evaluates job performance by "checking out" each individual on the "Performance Factors" listed in the Manual of Qualifications for Advancement in Rating. This check out is a qualifying type of measure, a hurdle, a prerequisite to taking a test for advancement. The Air Force requires a check out for each individual on the job training standards, completion of OJT program, success in training phase tests and supervisor's recommendation before an airman is eligible to take a Specialty Knowledge Test.

Conclusions:

1. Each test outline should be so developed that areas which can be tested by performance tests will be clearly identified.

2. Each test outline should clearly identify those elements of the job which can be demonstrated by performance, such as assemble, disassemble, trace, replace, etc.

Recommendations:

1. The Air Force furnish the Army with the Job Training Standards for the Air Force specialties to assist Army personnel in test outline development

and the identification of areas to which performance measures might be applied.

2. The Army continue its work on the development of Performance Check Lists.

Problem 4: "Test Outlines for Occupations which are composed of Duty Positions or Job Requirements which are heterogeneous - apparently unrelated to each other in many respects."

Discussion:

There are occupations in each of the services which are composed of duty positions or job requirements which are apparently unrelated to each other. These occupations are frequently relatively new or are older occupations which are beginning to be revised because of new developments in such things as equipment, doctrine, training, etc. Evaluating personnel in such occupations is usually a requirement because information is needed in order to effect personnel actions such as award of proficiency pay, promotion, and determination of job qualification. Developing tests for such occupations is frequently a problem because adequate coverage of heterogeneous job requirements in one test is difficult. Interpretation of test results also presents a problem since it cannot always be assumed that personnel are trained, assigned, and utilized in all aspects of the job. Even though those concerned with test development may feel that the structure of the job needs changing, it must be recognized that this is not the mission of these individuals. Their mission, rather, is to develop if possible a test which will do the best job of sampling the duties, skills, and knowledges required in the occupation. It may be that, in the opinion of those charged with test development, it is not possible to develop a test which will assist in evaluating personnel in the occupation. If this is the case, then they should so state their opinion with reasons for it.

Conclusions:

1. Each test outline must reflect accurately the duties, skills, and knowledges as determined from official sources.

2. Develop the test outline in such a way that areas of the job which are common to the various requirements can be identified.

3. Assign the majority of the items or as many as possible to the areas which are common. Assign equal weights or nearly so to each area which is unique.

Recommendation:

Construct a theoretical piece of equipment or a theoretical situation embodying the common elements of the job and develop items based upon it.

Actions Requested:

1. The Performance Requirements Survey referred to in the Proceedings of the Second Annual Conference of the Military Testing Association was distributed to participants in Committee 5 last year. Most participants in this Committee received copies of this survey this year. It was requested that the Survey be utilized in the various services as an aid to test outline development and that the Chairman be given reports concerning its value. These reports will be made available to other members of the Committee.

2. It was requested by the Chairman that any new ideas concerning test outline development be made known so that all members of the Committee can benefit from such reports.

3. The Chairman urged all members to exchange any articles written on the development of test outlines.

Report of Discussion Committee #6

Development of Test Items
(Training of Item Writers)

Participants	Affiliation
Mr. R. V. Baker, Chairman	Pers Lab, USAF
Mr. W. W. Wance	USAEEC
Lt M. W. Hallock	USCGI
CHBOSN S. P. Quinones, Jr.	USCGI
LCDR J. H. Miller	USNEC
LCDR R. E. Malmfeldt	USNEC
Mr. J. Crediford	USNEC
Mr. C. A. Hudson	USNEC
Lt M. A. Bachtel	USNEC
Mr. G. P. Cronkhite	ECI
Capt D. H. Taylor	Pers Lab, USAF
Mrs. V. A. Tribble	Pers Lab, USAF
Mr. S. W. Fotis	Pers Lab, USAF
Mr. C. Burkholder	LMTC, USAF
Mr. C. B. Chaffin	LMTC, USAF
Mr. R. L. Mitchell	LMTC, USAF
Capt J. A. Nolan	LMTC, USAF
Mr. B. E. Williams	LMTC, USAF
MSgt D. G. Carmier	USAMFSS
MSgt N. Gracernin	USAMFSS
Mr. V. S. Bowser	USA Ord Sch
Mr. N. Yannuzzi	USA Ord Cmd
Mr. D. Manch	ENG, DA
Mr. S. Baurmash	USAAA
Capt F. H. Blickle	USA Info Sch
Lt R. A. Ball	TAGSUSA
Lt V. Christensen	USA Arty & Msl Sch
Lt J. R. Barrett	USA (N Sch Music)
Mr. T. M. Ansbro	USA QM Sch
Capt D. C. Martin	USA Arm Sch

The topic for discussion in this committee grew out of a combination of two problems which were closely interrelated: (1) The training of item writers, and (2) the development of test items. The committee agreed that the distance from a properly trained item writer to an adequate multiple-choice test item should be a relatively short one.

As a point of departure in committee discussion, it was the concensus of opinion that the most important area for discussion in this committee was the careful selection of men to be trained as item writers. It was agreed that the importance of item writing cannot be overemphasized. If this first premise is true, the item writer should assume relative importance. Therefore, there should be rank (military) for the item writer commensurate with the importance of his mission.

It was agreed that the training of an item writer is never static. Item writer training is a fluid, constant development. It must be kept current and the item writer must be as current, at least, as the field in which he is writing. Selection and choice vary with the branch of the service. Training techniques also vary.

There were very few common problems and there is obviously no best system. One of the problems which comes nearer being common than most is the problem of obsolescence of test items. The various services represented agreed that this is a problem common to all the services. However, discussion brought out the fact that each service has different methods for solving this problem.

There was much discussion concerning "field" doctrine and "school" doctrine in all the services represented. The Army representative discussed the feasibility of items based on local law or "policy."

It was generally agreed that the item writer should be a highly motivated individual who is technically competent, academically oriented, of high I. Q., possesses inherent good judgment, has obligated service time, and should be married.

Recommendations:

The committee recommended that plans should be developed for the creation of an item writers' handbook which would be common to and available to all

the services. This recommendation was well received by the MTA in the general meeting. Captain Greer of the USNEC volunteered to host any meeting of this group who planned to work on the development of this handbook. It was agreed in committee that plans and discussion for the creation and development of this handbook would continue throughout the year and the recommendation would be placed high on the agenda for the next MTA Conference.

ANNEX 1, REPORT OF COMMITTEE #6

DEVELOPMENT OF TEST ITEMS (TRAINING OF ITEM WRITERS)

1. What steps should be taken by the parent command in selecting men for training as item writers?

2. Should the "prestige" factor be emphasized as a motivational factor in this selection?

3. Importance of early orientation of item writers and test psychologists to each other.

4. How may rapport be best established?

5. Training techniques and methods for training item writers.

6. Discuss the idea that the training of item writers is a two-way learning process. Unless the test psychologist is receiving and giving at the same time the item writer trainee is receiving (and giving), there will be difficulty in reaching a synthesis of ideas.

7. There seems to be general agreement that the best form of multiple choice items for use in job proficiency tests is the item with four distracters. It is not felt that the discussions or conclusions will vary from this general agreement.

8. Should the rule of the central problem in the stem be rigidly adhered to?

9. Should item statistics be the sole determinant of items considered for reuse in a revised test?

10. What ultimate value has the test item distracter?

11. What should be the motivational factors in group and individual productivity of test items?

12. Should the item writer be concerned with the type of man who enters into his career field?

Report of Discussion Committee #8

Automation in Processing Evaluation Instruments

Participants	Affiliation
Capt E. Madril, Chairman	Pers Lab, USAF
Lt E. Glenn	USNEC
Ens C. Marzallo	USNEC
Mr. R. L. Guy	USNEC
Mr. B. Allison	USACMLC Sch
MSgt B. Warner	Pers Lab, USAF
Capt D. Schmitt	USA Sig Tng Cen
Mr. T. E. Chandler	USA Sig Tng Cen
Mr. C. Rudinski	USA Sig C Sch

The discussions participated in by members of the committee centered on three main topics. These were: (1) Automation of computer input information, (2) treatment of small sample test data, and (3) psychologists and electronic computers.

Topic 1 - Automation of Computer Input Information: The conversion of source information into acceptable computer input media is a costly and time consuming phase of a test processing operation which has been automated. This is particularly true when data to be processed through electronic computers is generated at dispersed locations and then forwarded to a centralized activity for treatment. Such methods assume the collection of large volumes of information which must through some means be reduced to a form acceptable by the electronic computer system available to the activity concerned.

Three methods have been used broadly and considerable experience has been acquired. These are: (1) The preparation of data, using key punch devices for the production of machine record cards and/or paper tape; (2) the use of manual perforation of pre-scored machine record cards, and (3) the use of electro mechanical mark sense equipment for the preparation of machine record cards. Participants in the discussion agreed that each of the three methods is subject to limitations which can readily be overcome

by equipment which can be built and for which the state of present day technology is ready. However, the engineering and development costs at present are excessive. Interested members of the committee will continue to be on the alert for the development of equipment which may become commercially available and which will meet the needs of the services in reducing test data to suitable computer input media at reasonable cost. Continued consideration will be given but not limited to optical and electronic sensing equipment.

Topic 2 - Treatment of Small Sample Test Data: It was generally concluded by the members of the committee that we will have to continue to contend with small samples when administering tests on an operational basis. Statisticians and psychologists will have to resort to intuition, experience and whatever knowledge they possess about the tests in order to develop applicable norms. Of course the tools available to the statisticians commonly referred to as small sample statistical tools, will have to be used judiciously.

Topic 3 - Psychologists and Electronic Computers: The committee members concluded that psychologists will in time depend heavily on electronic computers for the treatment of their data and will find it much to their advantage to become at least conversant with people skilled in programming and operating such equipment as they have heretofore found it advantageous to be conversant with statistical clerks working with desk-type calculators and nomographs.

ACTIVITIES REPRESENTED AT THE CONFERENCE

Activity	Abbreviated Title
U. S. Army Enlisted Evaluation Center Fort Benjamin Harrison Indianapolis 49, Indiana	USAEEC
The Adjutant General's School, US Army Fort Benjamin Harrison Indianapolis 16, Indiana	TAGSUSA
U. S. Coast Guard Institute Groton, Connecticut	USCG Inst
U. S. Naval Examining Center Great Lakes, Illinois	USNEC
Directorage of Personnel Planning, Hq USAF Bldg T-8 Washington 25 DC	HQ USAF
Extension Course Institute Gunter AFB, Alabama	ECI
Behavioral Science Laboratory Aerospace Medical Division Wright-Patterson AFB, Ohio	B Sci Lab
Bureau of Naval Personnel Department of the Navy Washington 25 DC	BUPERS
R&D Command Department of Army Washington 25 DC	R&D, DA
USA Security Agency Training Command & School Ft Devens, Massachusetts	USASATC&S

Activity	Abbreviated Title
Hq USA Security Agency Arlington Hall Station Arlington 12, Virginia	USASA
Medical Field Service School Brooke Army Medical Center Ft Sam Houston, Texas	USAMFSS
Chief of Engineers Department of Army Washington DC	ENG DA
USA Engineer School Ft Belvoir, Virginia	USA ENG Sch
USA Audit Agency 2d & T Streets, N. W. Washington 25 DC	USAAA
Surgeon General Department of the Army Washington 25 DC	OTSG-DA
USA Provost Marshal General's School Ft Gordon, Georgia	USA PMG Sch
USA Ordnance School Aberdeen Proving Ground, Md	USA Ord Sch
USA Ordnance Training Command Aberdeen Proving Ground, Md	USA Ord TC
USA Ordnance Guided Missile School Redstone Arsenal, Alabama	USA Ord GM Sch
USA Artillery and Missile School Ft Sill, Oklahoma	USA Arty & Msl Sch
USA Chemical Corps School Ft McClellan, Alabama	USACMLC Sch

Activity	Abbreviated Title
Personnel Laboratory Lackland AFB, Texas	Pers Lab, USAF
Lackland Military Training Center Lackland AFB, Texas	LMTC
USA Infantry School Ft Benning, Georgia	USA Inf Sch
U. S. Army Armor School Ft Knox, Kentucky	USA Arm Sch
USA Quartermaster School Ft Lee, Virginia	USA QM Sch
U. S. Army Element (Naval School of Music) U. S. Naval Station Washington DC	USA (N Sch Music)
USA Signal Training Center Ft Gordon, Georgia	USA Sig Tng Cen
Air Force Systems Command Andrews AFB, Md	AFSC
U. S. Army Transportation School Ft Eustis, Virginia	USA Trans Sch
USA Signal School Ft Monmouth, New Jersey	USA Sig C Sch
USA Army Information School Ft Slocum, New York	USA Info Sch
USA Air Defense School Ft Bliss, Texas	USA AD Sch

LIST OF CONFEREES

Name	Affiliation
Ablett, W. J., Maj	Pers Lab, USAF
Albritton, K. D., Lt	USCG Inst
Allison, W. B., Jr.	USACML Sch
Allred, N. J.	USASA
Anderson, C. C., Lt Col	Hq USAF
Ansbro, T. M.	USA QM Sch
Bachtel, Mary, Lt	USNEC
Baker, R. V.	Pers Lab, USAF
Ball, R. A., Lt	TAGSUSA
Barra, A. P.	LMTC
Barrett, J. R., Lt	USA (N Sch Music)
Baurmash, S. S.	USAAA
Bercaw, R. V., CHSPCK	USCG Inst
Blickle, F. H., Capt	USA Info Sch
Bower, C. L., A1C	Pers Lab, USAF
Bowser, V. S.	USA Ord Sch
Boyd, J., Dr.	USA Ord GM Sch
Bridges, C. F.	USAEEC
Briscoe, C. B., CDR	USNEC
Burkholder, C.	LMTC
Cantrell, G. K., Capt	Pers Lab, USAF
Carmier, D., MSgt	USAMFSS
Carp, A., Dr.	Pers Lab, USAF
Chaffin, J. B.	LMTC
Chandler, T. E.	USA Sig Tng Cen
Christensen, V., Lt	USA Arty & Msl Sch
Cowan, G. E., Capt	Pers Lab, USAF
Crediford, J.	USNEC
Cronkhite, G. P.	ECI
Curry, R. R., Capt	USCG Inst
Davis, T. A., TSgt	Pers Lab, USAF
Dukerschein, E. A.	USNEC
Felthaus, J. J., Maj	Pers Lab, USAF
Foley, B. J.	USASATC&S
Fotis, S. W.	Pers Lab, USAF
Glenn, E. R., Lt	USNEC
Gracenin, N., MSgt	USAMFSS
Granger, R. H., CWO	USA Ord GM Sch

Name	Affiliation
Greer, H. H., Jr., Capt	USNEC
Guy, R. L.	USNEC
Hallock, M. W., Lt	USCG Inst
Haradon, C. B.	Pers Lab, USAF
Harding, F. D., Dr.	Pers Lab, USAF
Henley, S., Capt	OTSG-DA
Herring, T. G., TSgt	LMTC
Hodges, C. W.	BUPERS
Holdrege, F. E., Col	Pers Lab, USAF
Hudson, C. A.	USNEC
Hyman, H. D.	LMTC
Johnston, L. W.	USNEC
Jones, J. D.	LMTC
Judy, C. J., Dr.	Pers Lab, USAF
Kielman, R. W., Lt	USAMFSS
Knauf, A. S., Lt Col	Pers Lab, USAF
Legere, C. L. J.	USASATC&S
Long, G. S.	Pers Lab, USAF
Louis, E., Maj	TAGSUSA
Lutz, C.	Pers Lab, USAF
Macaluso, C. J.	USNEC
Madden, H. L.	Pers Lab, USAF
Madril, E., Capt	Pers Lab, USAF
Malmfeldt, R. E., LCDR	USNEC
Manch, D.	ENG, DA
Martin, D. C., Capt	USA Arm Sch
Marzallo, C., Ens	USNEC
McGraw, J. F., Capt	USA Inf Sch
McLanathan, F. L., Maj	Hq USAF
Mestan, A. A., Capt	USA AD Sch
Miller, E. E., Col	USAEEC
Mitchell, R. L.	LMTC
Morsh, J. E., Dr.	Pers Lab, USAF
Naranjo, J. A., Jr.	LMTC
Newman, I. J.	Pers Lab, USAF
Nolan, J. A., Capt	LMTC
Parris, H. L., Col	B Sci Lab
Partington, J. E.	USAEEC
Pearce, C. J.	Pers Lab, USAF

Name	Affiliation
Quinones, S. P., Jr., CHBOSN	USCG Inst
Ramras, E. M.	BUPERS
Rogan, M. R., CDR	USNEC
Rose, L.	USA Trans Sch
Rowland, R. O., Capt	USA PMG Sch
Rudinski, C.	USA Sig C Sch
Russell, S. B., LCDR	USCG Inst
Sait, E. M.	USAEEC
Schmitt, D., Capt	USA Sig Tng Cen
Schutter, G. I.	USNEC
Shafer, J., Maj	AFSC
Shelton, W. E., Col	R&D, DA
Smith, C. D., Maj	Pers Lab, USAF
Smith, C. W.	LMTC
Taylor, D. H., Capt	Pers Lab, USAF
Teller, J. D., Dr.	Hq USAF
Tinkle, J. W.	Pers Lab, USAF
Tribble, Virginia A.	Pers Lab, USAF
Vannest, L. D.	Pers Lab, USAF
Vitola, B. M.	Pers Lab, USAF
Wance, W. W.	USAECC
Weisbrod, B.	USA ENG Sch
William, B. E.	LMTC
Yannuzzi, N.	USA Ord TC
York, Beverly K., Ens	USNEC
Zaccaria, M. A., Dr.	LMTC