DOCUMENT RESUME

ED 109 195 TM 004 645

AUTHOR	Foley, John P., Jr.
TITLE	Evaluating Maintenance Performance; An Analysis. Final Report.
INSTITUTION	Air Force Human Resources Lab., Wright-Patterson AFB, Ohio. Advanced Systems Div.
REPORT NO	AFHRL-TR-74-[redacted](1)
PUB DATE	Oct 74
NOTE	38p.; For related documents, see TM 004 444, 643 and 644
EDRS PRICE	MF-$0.76 HC-$1.95 PLUS POSTAGE
DESCRIPTORS	Criterion Referenced Tests; *Electronic Equipment; Electronics; Job Analysis; Military Personnel; *Performance Tests; *Task Analysis; *Task Performance; *Test Validity
IDENTIFIERS	*Air Force

ABSTRACT

Formal measuring devices used to ascertain the training success and promotion potential of maintenance personnel have been deficient in the area of job realism. In part, this may be due to reliance on paper and pencil testing rather than the use of effective job performance measurements. Data concerning job performance measurements applicable to Air Force electronic maintenance were collected, structured, and analyzed. As a result of this analysis, it was recommended that comprehensive exploratory and advanced development efforts concerning Air Force maintenance should be established and funded. These programs should systematically and comprehensively identify and solve problems concerning maintenance practice in the field and problems concerning the selection and training of maintenance personnel. A necessary first effort should be to gather and study hard data on how well maintenance men can perform the key tasks of their jobs. Also, based on the results of this analysis a contractual effort was initiated to develop criterion referenced JTPT for electronic maintenance tasks. This effort was followed by an attempt to develop both graphic and video symbolic substitutes of improved empirical validity. (Author/BJG)

AIR FORCE

HUMAN RESOURCES

AFHRL-TR-74-57(I)

EVALUATING MAINTENANCE PERFORMANCE: AN ANALYSIS

By

John P. Foley, Jr.

ADVANCED SYSTEMS DIVISION
Wright-Patterson Air Force Base, Ohio 45433

October 1974
Final Report for Period 1 January 1966 - 31 July 1974

Approved for public release; distribution unlimited.

LABORATORY

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

NOTICE

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This final report was submitted by Advanced Systems Division, Air Force Human Resources Laboratory, Wright-Patterson Air Force Base, Ohio 45433, under project 1710, with Hq Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base, Texas 78235.

This report has been reviewed and cleared for open publication and/or public release by the appropriate Office of Information (OI) in accordance with AFR 190-17 and DoDD 5230.9. There is no objection to unlimited distribution of this report to the public at large, or by DDC to the National Technical Information Service (NTIS).

This technical report has been reviewed and is approved.

GORDON A. ECKSTRAND, Director
Advanced Systems Division

Approved for publication.

HAROLD E. FISCHER, Colonel, USAF
Commander

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFHRL-TR-74-57 (I)	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) EVALUATING MAINTENANCE PERFORMANCE: AN ANALYSIS		5. TYPE OF REPORT & PERIOD COVERED Final 1 January 1966 – 31 July 1974
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) John P. Foley, Jr.		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Advanced Systems Division Air Force Human Resources Laboratory Wright-Patterson AFB, Ohio 45433		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62703F 17101006
11. CONTROLLING OFFICE NAME AND ADDRESS Hq Air Force Human Resources Laboratory (AFSC) Brooks Air Force Base, Texas 78235		12. REPORT DATE October 1974
		13. NUMBER OF PAGES 38
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This report consists of 4 volumes.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

symbolic substitute job tests
Job Task Performance Tests
criterion referenced tests
video testing
maintenance effectiveness
personnel selection
psychology - individual and group
measurement and evaluation technical training
measurement and evaluation maintenance training
measurement and evaluation electronics training
measurement and evaluation
vocational education

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Late in 1962, the writer prepared a paper entitled *Performance Testing: Testing for What is Real* (Foley, 1963). This was followed by a *Bibliography on Maintenance Personnel Performance Measurement* by Askren (1963) and work on a draft of an annotated bibliography by Porterfield in the same year. All of these efforts indicated that there was probably a lack of job realism in the formal measuring devices used to ascertain the training success and promotion potential of maintenance personnel. They also indicated that although a rather extensive technology had been developed for paper and pencil testing, no well-structured technology or guidance existed for the development and administration of job performance tests. These preliminary works and findings indicated a requirement for exploratory development concerning job performance tests for maintenance.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Item 20 (Continued)

To avoid duplication in such an exploratory development program, the approach was to tap the wealth of existing, but scattered, sources of available hard data concerning job performance measurements, to structure these data as they applied to the measurement of ability to perform electronic maintenance tasks, to analyze them in relation to current Air Force practice and to make recommendations for the development and tryout of effective job performance measurements for Air Force electronic maintenance.

Paper and pencil testing procedures are used almost exclusively for determining which personnel are selected for training, for determining student progress while in training and for determining the promotion eligibility of personnel assigned to field maintenance units. A number of studies are cited which indicate that low correlations were obtained by comparing job task performance tests to paper and pencil theory tests and to job knowledge tests. Several studies also are cited which indicate that the traditional theory content, found in most electronic maintenance training programs, does not contribute a great deal to the ability to perform electronic maintenance tasks. A full application of the modern technology for technical training development would solve the course content problem. This technology requires a systems approach to training program development in which training objectives are based on a complete job task identification and analysis. Criterion referenced Job Task Performance Tests (JTPT) are required to determine if training objectives are achieved. However, a serious gap remains in this technology since adequate guidance is not available for the development of JTPT.

As a result of this analysis, it was recommended that comprehensive exploratory and advanced development efforts concerning Air Force maintenance should be established and funded. These programs should systematically and comprehensively identify and solve problems concerning maintenance practice in the field and problems concerning the selection and training of maintenance personnel. A necessary first effort should be to gather and study hard data on how well maintenance men can perform the key tasks of their jobs. Also, based on the results of this analysis a contractual effort was initiated to develop criterion referenced JTPT for electronic maintenance tasks. This effort was followed by an attempt to develop both graphic and video symbolic substitutes of improved empirical validity. The results of these efforts are reported in Volumes II, III, and IV of this series of documents.

SUMMARY

Problem

The application of the technology of programmed instruction in Air Training Command resulted in a number of lists of so-called measurable behavioral objectives. After reviewing several of those lists in 1962, the writer felt that the behaviors listed were verbal behaviors that could be conveniently measured by paper and pencil objective type tests, but that the listed behaviors were not necessarily the behaviors of the jobs for which the students were being trained. He felt that this was especially true for the key behaviors of maintenance jobs. Late in 1962, the author prepared a paper entitled *Performance Testing: Testing for What is Real* (Foley, 1963). This was followed by a *Bibliography on Maintenance Personnel Performance Measurement* by Askren (1963) and work on a draft (not published) of an annotated bibliography by Porterfield in the same year. All of these efforts indicated that there was probably a lack of job realism in the formal measuring devices used to ascertain the training success and promotion potential of maintenance personnel. They also indicated that although a rather extensive technology had been developed for paper and pencil testing, no well-structured technology or guidance existed for the development and administration of job performance tests. These preliminary works and findings indicated a requirement for exploratory development concerning job performance tests for maintenance. These works also indicated that there had been more reports concerning performance measures for electronic maintenance than for mechanical maintenance. This emphasis is reflected in the following approach.

Approach

To avoid duplication in such an exploratory development program, the approach was to tap the wealth of existing, but scattered, sources of available hard data concerning job performance measurements, to structure these data as they applied to the measurement of ability to perform electronic maintenance tasks, to analyze them in relation to current Air Force practice and to make recommendations for the development and tryout of effective job performance measurements for Air Force electronic maintenance. The most productive source of information directly applicable to proficiency measurements for electronic technology courses was the Defense Documentation Center. Approximately eighty reports were found concerning different aspects of proficiency measurement for electronic courses. Many of the same reports were also referenced in the *Psychological Abstracts.* Applicable material, as well as basic references, were obtained from several textbooks. Although the *Educational Index* was a valuable informational source for general vocational and industrial education, it contained very little reference materials applicable to measurement in these areas.

The literature on vocational education indicates that considerable emphasis has been placed on the development of objectives and materials for highly job relevant courses. However, very little attention has been given to the validity of measurement procedures for ascertaining that these job objectives have been achieved.

Results

Paper and pencil testing procedures are used almost exclusively for determining which personnel are selected for training, for determining student progress while in training and for determining the promotion eligibility of personnel assigned to field maintenance units. This analysis of reported research indicates that these measurement procedures do not insure that these personnel can perform the electronic maintenance tasks generated by the electronic hardware they are assigned to maintain. Most of the paper and pencil tests used in training and promotion concern theory and job knowledge. Studies are cited which indicate that extremely low correlations were obtained by comparing job task performance tests to paper and pencil theory tests and to job knowledge tests. Several studies are also cited which indicate that the traditional theory content, found in most electronic maintenance training programs, does not contribute a great deal to the ability of an individual to perform electronic maintenance tasks. However, it is widely contended that a knowledge of theory is a necessary prerequisite to the successful performance of such tasks and this contention is deeply imbedded in the electronic maintenance culture. A full application of the modern technology for technical training development would solve the course content problem. This technology requires a systems approach to training program development in which training objectives are based on a complete job task identification and analysis. Criterion referenced Job Task Performance Tests (JTPT) are required to determine if training objectives are achieved. However, a serious gap remains in this technology, since adequate guidance is not available for the development of JTPT. When administered, JTPT have been expensive in terms of equipment, time and administrative personnel. JTPT should be utilized in spite of their high cost if they are the only empirically valid tests available. If empirically valid symbolic substitute tests could be produced, they certainly should be used. After a study of the symbolic substitute Multiple – Alternative Symbolic Troubleshooting Test (MAST) and Tab Tests, it was hypothesized that the empirical validity of symbolic substitute of this type might be improved by increasing their realism.

Conclusions

Based on the results of this analysis a contractual effort was initiated to develop criterion referenced JTPT for electronic maintenance. This effort was followed by an attempt to develop both graphic and video symbolic substitutes of improved empirical validity. The results of these efforts are reported in Volumes II, III, and IV of this series of documents (see Preface). The preparation of a guide for developing JTPT is recommended. Comprehensive exploratory and advanced development efforts should be established and adequately funded concerning Air Force maintenance. These programs should systematically identify and solve problems concerning maintenance practice in the field and problems concerning the selection and training of maintenance personnel. A necessary first effort should be to gather and study hard data on how well maintenance men can perform the key tasks of their jobs.

PREFACE

This document represents a portion of the exploratory development program of the Advanced Systems Division, Air Force Human Resources Laboratory, Wright-Patterson Air Force Base, Ohio. This document is the first of four volumes to be published concerning the evaluation of maintenance performance. Volumes II, III, and IV are:

1. *Evaluating Maintenance Performance: The Development and Tryout of Criterion Referenced Job Task Performance Tests for Electronic Maintenance.* AFHRL-TR-74-57(II), Part I, in press.

2. *Evaluating Maintenance Performance: Test Administrator's Manual and Test Subject's Instructions for Criterion Referenced Job Task Performance Tests for Electronic Maintenance.* AFHRL-TR-74-57(II), Part II, in press.

3. *Evaluating Maintenance Performance: The Development of Graphic Symbolic Substitutes for Criterion Referenced Job Task Performance Tests of Electronic Maintenance.* AFHRL-TR-74-57(III), in press.

4. *Evaluating Maintenance Performance: A Video Approach to Symbolic Testing of Electronic Maintenance Tasks.* AFHRL-TR-74-57(IV), in press.

The preparation of these volumes has been documented under Task 171010, Evaluating the performance of Air Force Operators and Technicians of Project 1710, Training for Advanced Air Force Systems. The effort represented by this volume was identified as work unit 17101006. The author was the task scientist. Dr. Ross L. Morgan was the project scientist.

The author wishes to express his appreciation to Dr. William B. Askren, Advanced Systems Division, and Mr. James Porterfield, Aerospace Medical Research Laboratories, for their original biographical works which were used by the author for the original identification of sources for the author's dissertation upon which this analysis is based. He also wishes to express his appreciation for the guidance given by his dissertation committee: Dr. Dennis H. Price, chairman; Dr. Howard B. Lyman and Dr. Jack Corle — all of the University of Cincinnati. The final product was influenced by many discussions with Dr. Ross L. Morgan and Dr. Gordon A. Eckstrand of the Advanced Systems Division.

TABLE OF CONTENTS

EVALUATING MAINTENANCE PERFORMANCE: AN ANALYSIS

I. BACKGROUND

This report (AFHRL-TR-74-57 (I)) is the first of a series of four volumes pertaining to the measurement of job proficiency. It presents the background and analyses that resulted in the funding of a series of exploratory development contracts starting in 1969. These contractual efforts initially concerned the development and tryout of criterion referenced Job Task Performance Tests and later the development and tryout of both graphic and video symbolic substitute tests. The results of these follow-on efforts are reported in Volumes II, III and IV of AFHRL-TR-74-57. Several in-house efforts and events prior to 1969 were reflected in the writer's preparation of the early work statements for these contractual efforts.

Late in 1962, the writer prepared a paper entitled *Performance Testing: Testing for What is Real* (Foley, 1963). The motivation for this paper was the review of several lists of measurable behavioral objectives which had been generated in Air Training Command (ATC) as the basis for preparing programmed instruction packages. The writer felt that the behaviors listed were verbal behaviors that could be conveniently measured by paper and pencil objective type tests, but that the listed behaviors were not necessarily the behaviors of the jobs for which the students were being trained. It was the writer's contention that a large portion of the desired job task behaviors involved complex perceptual-motor skills. He felt that the only way such behaviors could be adequately measured was by performance tests in which each test subject was required to demonstrate that he could in fact perform each key task of his job. As a follow on to this first memorandum report, early in 1963, Dr. William B. Askren prepared a *Bibliography on Maintenance Personnel Performance Measurement* (Askren, 1963). During the summer of 1963, Mr. James Porterfield, a graduate student from Kansas State University, expanded and annotated the Askren bibliography. This draft bibliography was never published.

In 1964, Eckstrand published a paper concerning the status of the technology of training in which he discussed measurement in the context of a system's approach to training. He suggested the use of criterion referenced measures as appropriate for measuring student achievement of carefully derived training objectives, stated in terms of what a graduate should do. He also discussed the difference between criterion referenced and norm referenced tests, and referenced Glaser and Klaus (1962) as his major source for this distinction.

In the 1965 to 1967 time period, the writer prepared a review of the literature for his doctoral dissertation concerning measurement practices of Post-High School Vocational Electronic Courses (Foley, 1967a). These courses are very similar in content to the Air Force's electronic maintenance courses. This review of the literature made use of the previous biographical works of Askren and Porterfield as well as sources discussed later. The writer's thinking expressed in this dissertation was influenced by his many discussions with Dr. Ross L. Morgan and Dr. Gordon A. Eckstrand Many of the materials contained in this volume appear in the dissertation. The dissertation materials have been modified to apply more directly to Air Force measurement problems.

In 1966, the writer monitored and participated in a field survey of electronic maintenance in the Air Force (Folley & Elliott, 1967). The survey information gathered during this study indicated that maintenance practices in the Air Force could be greatly improved. The authors were aware of the softness of data, gathered by interviews and observation, but as the survey indicated, no recorded hard data existed concerning how well each maintenance man could perform his various job tasks. The generation of such performance data would require the administration of objective job performance tests which were not available.

Further motivation and input for preparing the original purchase requests for the follow-on performance measurement efforts (reported in Volumes II, III and IV) were provided by two conferences on measurement which were supported by the Advanced Systems Division, Air Force Human Resources Laboratory, Wright-Patterson Air Force Base, Ohio. In March 1966, the writer attended a conference on the *Assessment of Complex Operator Performance*. This conference was arranged and reported by Dr. W. Dean Chiles (1967), but was chaired by Dr. Arthur W. Melton of the University of Michigan. In October 1968, the writer participated in a conference chaired by Mr. Melvin T. Snyder (Snyder, Kincaid, & Potempa, 1969). This conference concerned human factors testing. Both of these conferences indicated to the writer that there was an urgent requirement for the development of more effective measuring devices for ascertaining how well Air Force personnel could perform the maintenance job tasks generated by existing and future hardware systems.

The purpose of the work reported in this volume was to tap the wealth of existing, but scattered, sources of already available information concerning job performance measurements, to structure this information as it applied to the measurement of ability to perform electronic maintenance tasks, to analyze it in relation to current Air Force practice and to make recommendations for the development and tryout of effective job performance measurements for Air Force electronic maintenance. One primary purpose of a project such as this is to eliminate as far as possible duplication of work that has already been adequately accomplished. As mentioned previously, a major portion of the information had been identified as part of the writer's doctoral dissertation (Foley, 1967a). But in that document the information had been analyzed and reported for a somewhat different purpose.

The most productive source of information for these efforts has been the Defense Documentation Center (DDC). Approximately eighty reports were found concerning different aspects of proficiency measurement for electronic courses. Many of the same reports were also referenced in the *Psychological Abstracts*. Applicable materials, as well as basic references, were also obtained from texts by Super and Crites (1962), Thorndike (1949), Travers (1950, 1955), and Micheels and Karnes (1950). Although the *Educational Index* was a valuable informational source for general vocational and industrial education, it contained very little reference materials applicable to measurement in these areas. The literature on vocational education indicates that considerable emphasis has been placed on the development of objectives and materials for highly job-relevant courses. However, very little attention has been given to the development of valid measurement procedures for ascertaining that these job objectives have been achieved.

II. THE CRITERION PROBLEM

In most cases when we wish to determine whether an individual or group of individuals can perform behaviors in a given environment we cannot simulate all the desired conditions. If we can produce a measuring device that actually measures the ability to perform the desired behaviors under all the desired conditions, we have an ultimate criterion measure. But the fact that we usually cannot develop such a device forces us to settle for a secondary criterion measure which is, at best, somewhat different than the ultimate criterion measure. This difference between the real world and the simulation of the real world for testing purposes is called the criterion problem.

A common example of such a criterion problem presents itself when we attempt to measure an individual's ability to drive automobiles. To measure such ability completely we would have to devise a test that would measure his ability to perform all driving tasks of all automobiles, on all types of roads, in all traffic conditions, under all types of weather conditions, whether he is being observed or not. It is obvious that it would be virtually impossible to meet all of these conditions under practical testing conditions. We, therefore, settle for a less rigorous test criterion. We assume that he can drive any automobile adequately, if he demonstrates in a performance test that he can perform most driving tasks in one automobile, in normal traffic, while being observed.

But many times, it is inconvenient and considered too costly to administer even such a driver performance test and an attempt is made to develop a paper and pencil test which will determine that an individual can drive adequately. But such a test cannot be considered to be valid substitute unless a high empirical relationship to the criterion measure can be demonstrated. In the practical world of test development, the driver performance test would be considered an adequate near ultimate criterion test for validation of such a paper and pencil substitute. Many times such a paper and pencil test is used without being validated against such a near ultimate criterion test. The use of such an unvalidated test is an extremely dangerous practice, since it is assumed by most users that it measures an individual's ability to drive when in fact we are not sure what it is measuring.

This criterion problem has long plagued measurement theorists and practitioners as well as curriculum researchers. The use of job tasks, and performance examinations based on these tasks as near ultimate criteria for evaluation of selection devices, was first emphasized as a result of the work of Army and Navy measurement psychologists during World War II. In 1946, Jenkins discussed the problem in light of the experiences of Navy psychologists in an article titled "Validity for What?"

> Psychologists in general tended to accept the tacit assumption that criteria were either given of God or just to be found lying about....The novice of 1940, searching through many textbooks and much journal literature would have been led to conclude that expediency dictated the choice of criteria and that the convenient availability of a criterion was more important that its adequacy.

In 1964, Wallace presented a paper at the annual convention of the American Psychological Association in which he indicated that much of what Jenkins said in 1946 was still true (Wallace, 1965a).

Super and Crites (1962), and Thorndike (1949) have discussed this same problem in light of their Army experience. They have indicated that, although course success was a readily available validation criterion for prediction, success in the course often had little influence on job success. One of Thorndike's examples from the Army Air Force (AAF) is as follows:

> Early in WW II, gunnery schools in the AAF placed a good deal of emphasis in their program on learning the nomenclature of machine guns and turrets and on being able to express verbally the operation of this equipment. Using grades obtained in this type of program as a criterion, it was possible to develop a battery of verbal tests which gave a substantial prediction of those grades. Actual combat gunnery, however, presented no special verbal demands, and the type of memoriter training referred to above was eventually done away in favor of more and more training in the actual assembly, maintenance and firing of the guns. The nature of grades in gunnery school changed correspondingly. With this change in the criterion, the validities of verbal selection tests largely disappeared. Though there is no evidence in this case that the selection of gunners in terms of verbal abilities would have done any actual *harm* to the final output, it certainly would have been wasteful and ineffective. Selection would have been based on the *irrelevant* variance in the partial criterion of training grades (Thorndike, 1949, p. 126).

Thorndike goes on to clarify the relationship among predictive tests, training and job performance.

> There are many other instances, both in and out of the military situation, which *academic grades* are used as criterion measures because they are conveniently available and because they appear to possess a rather satisfactory degree of reliability. These grades can generally be predicted with fair success, and the research worker *may be lulled into a sense of satisfaction and accomplishment by this success in predicting them.* Sometimes, of course, his satisfaction may be justified, because the nature of the training is based on essentially the same attributes as performance in the job. The possibility is always a real one, however, that the portion of academic achievement that we predict with our tests is not the part that is relevant to later success on the job. Whether the criterion is grades or some other type of record, we must always examine it critically to judge whether the aspects of the criterion which we predict will be relevant for the ultimate goal (Thorndike, 1949, pp. 126-127).

Since an electronics training course has as its ultimate objective job performance, the ultimate criterion against which the course content and course measurement should be validated is the job for which the student is being prepared. Wallace (1965a) very succinctly expressed the criterion problem as it applies to electronics personnel.

> All of this is prelude to my main thesis which is in no sense revolutionary, original, or controversial. I state it because it is honored in the breach. It is that the nature of our proficiency measures determines how we select, classify, train, maintain and assess our human resources. If the measures are largely irrelevant to the jobs we want done, we will select the wrong men, classify them incorrectly, and train them wrong. This is true because these proficiency measures are, or should be, the criteria against which we validate our selection and classification procedures and evaluate our training content and methodology or our supervisory techniques. Thus, if I use a test of advanced electronics theory as the proficiency measure for electronics maintenance and as the criterion against which to evaluate a test for selecting men to go into maintenance training, I will end up choosing a selection test which rejects men who are not well above average in both reading and arithmetic ability. In the process I might reject a great many who are outstanding in their ability to get their hands on a piece of machinery and make it work. I might also accept a number who (like myself) are so lacking in the simplest manipulative ability that their hands could have been cut off at the wrists at birth without seriously affecting their outputs. So, when I decided what proficiency measures to use, I also decided what kind of men I was going to put into training for the job.
>
> But it doesn't end there. For when I now approach the problem of how to train men to perform the tasks involved in the job, I must make decisions about what should be taught and what methods should be used in teaching it. The only way I have of reaching such decisions (except by divination which is, admittedly, not a rare procedure) is to measure and compare the performances achieved with various curricula and methodologies. So, in the case of the electronics maintenance course, I put in lots of reading about electronics theory and I produce graduates who can read and write electronics theory while their equipment deteriorates in hopeless inoperativeness (Wallace, 1965b, p. 4).

Frederiksen has expressed a similar view.

> It should be obvious that one cannot hope to provide appropriate methods for evaluating instructional outcomes unless he has a clear idea of what those outcomes should be. Yet evaluations of instructional programs are sometimes made in which criterion measures are chosen because of their easy availability rather than through a careful study of what the students are supposed to be able to do as a consequence of training (Frederiksen, 1962, p. 324).

The problem is made more complex by the fact that a given course often is used to train many individuals to work in a variety of jobs in the electronic field. To carry the logic of an ultimate criterion to its conclusion, each of the jobs is an ultimate criterion. To include all of these in a course or its measurement program would be impossible. This does not mean, however, that we should ignore the problem of preparing people for specific job tasks that are common or of measuring their proficiency in these tasks. We can examine the electronic jobs as they exist today and determine those tasks which are found in all or practically all electronic jobs. We can even predict with some degree of certainty the tasks of future jobs.

Thorndike (1949, p. 121) differentiates three categories of criteria: ultimate, intermediate, and immediate. He indicates that "A...really complete ultimate criterion is multiple and complex in almost every case." He also suggests that it is possible to agree on certain final job behaviors as being the closest approximation to the ultimate goal. Frederiksen makes the following applicable statement:

> The objective, presumably, is to get as close as is feasible to the ultimate criterion; but as has just been seen, when one gets too close to the real-life situation, control of the conditions for adequate observation is lost. Observations of real-life behavior is ordinarily not a suitable technique for measurement. The type of measure that is recommended for first consideration in a training evaluation study is the type which most closely approximates the real-life situation, that which, in this chapter, has been called eliciting lifelike behavior. If it is not feasible to wait for the behavior to happen in real life, then lifelike occasions can be provided for the behavior to occur in a test situation (Frederiksen, 1962, p. 334).

Admittedly, an examination made up of tasks removed from their actual job environment is not an ultimate criterion test. Under actual job situations, the graduate may have to perform these tasks in cramped quarters; under stresses of time, noise, heat, or cold; or with an excited boss interfering. These conditions of stress are usually not constant variables, but change from day to day and from hour to hour. The assumption usually has to be made that the individual can perform a task under conditions of stress, provided he can perform the same task *well* under normal conditions. A formal performance examination has its own set of stresses, which may not be the same as job stresses, but their presence may tend to offset the lack of job stresses. Formal, job task performance examinations are the closest usable simulation of the real maintenance jobs presently available. They are far better than no performance tests at all.

III. WHAT JOB ACTIVITIES SHOULD BE MEASURED

One of the problems which has confronted the measurement psychologist when he has tried to develop job task performance tests for electronics maintenance has been a lack of a usable classification of job activities which would be definitive enough to develop diagnostic test procedures. The problem has been to find categories of behavior which are meaningful to training and maintenance personnel and which lump similar or closely related behaviors into each category. In 1962, Bryan summarized and classified such activities. In 1966, Folley discussed the job tasks of the electronic technician based on many sources in an unpublished paper. From the Folley discussion, the writer developed the following list of tasks or activities:

1. Job activities or tasks associated with electronic equipment:
   - a. Performing equipment checkout procedures.
   - b. Adjusting, aligning, and calibrating.
   - c. Replacing of components — removing and replacing.
   - d. Isolating between-stage faults to particular state (or functional unit or physically replaceable unit).
   - e. Isolating within-stage faults to defective component (tube, solid state device, coil, capacitor, resistor, etc.).

2. Information gathering activities of electronic technician jobs

   a. Using oscilloscope to: (1) measure voltage, (2) measure frequency, (3) compare waveshape to waveshape standard, and (4) make high accuracy time base measurements and comparisons.

   b. Using the electronic voltmeter to measure various ranges of voltages in electronic equipment.

   c. Using the ohmmeter to measure direct-current resistance in electronic equipment.

   d. Using the signal generator to inject standard or know signals into equipment for test purposes.

   e. Using the tube checker to estimate quality of electron tubes.

   f. Using the transistor checker to estimate quality of transistors.

3. Handtool activities of electronic technician jobs:

   a. Using screwdrivers.

   b. Using pliers.

   c. Using diagonal cutters.

   d. Using soldering iron.

   e. Using soldering gun.

   f. Using wire strippers.

   g. Using machinist's wrenches.

   h. Using light machinist's hammer.

This list is not offered as an all-inclusive list of job activities but one that can be considered common to most of the jobs. It does not include administrative activities that are peripheral to actual maintenance activities. The five items in paragraph 1 of this classification concern tasks or activities associated with many actual items of equipment with which the technicians work. The six items in paragraph 2 are information gathering activities which all electronic technicians will encounter in the performance of the five equipment activities. From a logical point of view, it would appear that any training program, that has as its objective the preparation of students to work with electronic hardware, should include a large amount of practice in the use of such information gathering devices. Paragraph 3 includes most of the small tools used by electronic technicians.

This classification represents a convenient structure for the development of Job Task Performance Test (JTPT) procedures and accompanying scoring schemes which have broad applicability to specific electronics jobs and related training.

IV. CONTENT VALIDITY VS EMPIRICAL VALIDITY

The maintenance of Defense hardware systems is an absolutely necessary activity, but its cost is extremely high. Is this maintenance cost too high? No one really knows the answer. And important unknowns include the lack of extensive hard data on the efficiency and skill of the maintenance technician and the lack of extensive hard data on the effectiveness of maintenance training. By hard data are meant empirically valid data. The limited amount of published data available indicates that many electronic technicians are not adequately trained to perform the tasks of their assigned jobs and that they do not perform many of their job tasks with a high degree of proficiency (Anderson, 1962a; Folley & Elliott, 1967). Yet each Air Force electronic technician has a record of passing school marks and of a passing score on the Specialty Knowledge Tests (SKT). But there are no hard data available on how well he can perform the actual tasks of his job. How valid then are the school marks and the SKT scores? It depends on what type of validity we are talking about. The discussion that follows considers three types of validity: fact validity, content validity and empirical validity.

Face Validity means that the test looks as if it should be valid (Lyman, 1971).

Content Validity is somewhat similar to face validity, but is more systematic and sophisticated. Other names for it include logical validity, course validity, curricular validity and textbook validity. Both face and content validity are non-statistical (Lyman, 1971). This definition of content validity is applied to course content. Tests can be constructed to have job content validity. This subject is discussed later. To obtain good content validity for course testing some sort of test "blue print" is usually used to insure that all aspects of the subject matter or of the job are adequately sampled.

The tests given by teachers or professors in academic education situations usually have face validity. If they don't, the students complain. If carefully prepared, some of these tests have content (or course) validity. Many of the pencil and paper block examinations given in Air Training Command technical courses have good content validity, since wide use is made of test "blue prints" for test development. The same can be said for the carefully prepared Air Force paper and pencil SKT.

Empirical Validity is criterion-referenced validity and "is implied whenever no adjective is used to modify validity. This sort of validity is most important in *practical* situations. How well does the test measure what we want it to? Empirical validity gives us the answer by indicating how closely the tests relate to some criterion (i.e., to some standard of performance)" (Lyman, 1971). In some cases empirical validity is non-statistical; in others, it must be determined by statistical means.

If a test could be constructed which was an ultimate criterion measure for an electronic maintenance job, it would have perfect empirical validity. It would, also, have perfect job content validity because it would have to contain all of the job tasks performed in a true job environment. Although the tasks that an electronic maintenance technician must perform can be identified rather precisely, the environment in which each task is performed varies from performance to performance. The best we can hope for is a test environment which closely approximates a typical real-life situation. So just as there is no test which is a perfect ultimate criterion measure, there is no test that has perfect empirical validity. But if job tasks are carefully and systematically selected for a test battery, so as to be truly representative of the tasks of the job, and the test subjects are required to perform the selected tasks in an environment that reflects a typical job environment, it *can logically be assumed* that we are as close as feasible to the ultimate criterion. (The performance of each selected task is an individual performance test.) It can also be *assumed* that a battery made of such tests has practical empirical validity and job content validity. (The main purpose for presenting the classification of maintenance tasks in the previous section was to provide a frame of reference for the systematic selection of typical job tasks for the development of near-ultimate criterion tests.) The empirical validity of a job relevant test battery so developed has been obtained by the manner in which the battery has been constructed. This is a non-statistical type of validity. Such a battery of tests would be job criterion referenced. In this analysis such tests are called criterion referenced Job Task Performance Tests (JTPT). Since such a test battery comes as close as is feasible to the ultimate criterion, it can be used as a near ultimate criterion for the determination of the practical statistical empirical validity of other tests which do not require the test subjects to actually perform real job tasks. The next section of this analysis considers the statistical validity of a number of tests of the latter type, such as theory tests and job knowledge tests.

Theory tests and job knowledge tests may have some type of content validity. The content validity of an electronic theory test may be determined by systematically sampling the content of the academic subject matter called electronic theory. Such *course* content validity is not *job* content validity. A type of job content validity is obtained concerning job knowledge tests by asking questions about well-selected job tasks. However, in the job situation the technician is required to actually perform the job tasks whereas in the job knowledge test situation, the test subject is required to verbalize in some manner about job tasks. This is a "deed" to "word" relationship rather than the "deed" to "deed" relationship required of near ultimate criterion tests. To ascertain whether such a job knowledge test has high empirical validity, a high statistical correlation must be demonstrated with a test having the required "deed" to "deed" relationship.

Maintenance technical training courses in the Department of Defense and their civilian counterparts in vocational education are preparing their students for the world of work. How well the work objectives of these training programs are achieved can only be determined by the administration of tests having high empirical validity, based on students' and graduates' ability to perform the key tasks of their jobs. In the discussion of this report the criterion to be applied to all tests is "how well does each type of test measure each test subject's ability to perform the key tasks of the job for which he is being trained or the job to which he is assigned." The sections that follow present what has been found in the research literature concerning the empirical validity of course content and the empirical validity of the various types of tests now utilized, not only in maintenance training programs, but also, in field maintenance units.

V. THE VALIDITY OF JOB TASK PERFORMANCE TEST SUBSTITUTES

Since paper and pencil tests and other substitutes are much less difficult to administer and less time consuming, they are used extensively as measures of achievement. Examples of other substitute measures include symbolic equipment tests such as the Tab Test (Crowder, Morrison, & Demaree, 1954) and the Multiple-Alternative Symbolic Troubleshooting Test (MAST) (Grings, Rigney, Bond, & Summers, 1953) as well as peer ratings and supervisors' ratings. Travers (1955) describes various substitutes for job performance tests such as oral questioning, printed verbal tests (paper and pencil), and illustrated or pictorial tests.

Research evidence, however, gives a rather low rating to all of these substitutes. Table 1 shows correlations that have been obtained by comparing Job Task Performance Tests (JTPT) to theory tests, and to job knowledge tests. The latter two are paper and pencil tests. Table 1 also includes correlations with school marks. In most cases, school marks have been heavily weighted with the paper and pencil test scores. An examination of this table indicates that the correlations of JTPT scores with theory test scores are generally somewhat lower than with job knowledge tests.

Probably the most extensive study regarding substitutions of a paper and pencil achievement test for actual JTPT was made as part of Human Resources Research Office Project REPAIR (Brown, Zaynors, Bernstein, & Shoemaker, 1959). To check the effectiveness of the revised repair course against the traditional course, a proficiency battery made up of six separate tests was devised. The battery included four JTPT. JTPT were developed and individually administered in the areas of troubleshooting, test equipment, repair skills, and alignment. Two paper and pencil group tests were given on the use of maintenance manuals and the reading of schematic diagrams. A seventh achievement test was developed upon the insistence of school personnel that a paper and pencil test was required in lieu of the four JTPT. This test consisted of seventy-seven multiple choice items on topics of knowledge concerning troubleshooting, use of schematics, use of test equipment, operation checks, and alignment. As presented in Table 1, the correlation between this achievement test and the performance troubleshooting test was .40; with alignment test, .28; the repairs skills test, .19; with the test equipment, .29; with the paper and pencil manuals test, .51; and with the paper and pencil schematic test, .51. The authors of the report state that, as would be expected, the correlations are the highest with other paper and pencil tests; namely, the manuals tests and schematic tests. The authors also state that although the achievement tests correlate positively with most of the other tests in the battery, *the correlations are not high enough to justify the use of the achievement test as a substitute for the battery of JTPT.* This statement also would apply to all other tests presented in Table 1.

The Saupe (1955) study indicated somewhat higher correlations than the other studies presented. The correlation between a troubleshooting performance test and a job knowledge test was .55 and with school marks was .56. However, the troubleshooting problems were performed on a very simple radio receiver built specifically for the study on an open chassis. The test subjects also used two very simple test equipments — a voltohmmeter and signal generator. All of the other studies used actual military electronic hardware of greater complexity requiring more complex test equipment.

A more recent study by Finucane (1966) in the field of engine repair in the Army results in very similar conclusions. The results of a JTPT on various aspects of engine maintenance were compared with job knowledge tests concerned with the same engines. The correlation between the two types of tests was .31, which is within the same range as the findings relative to the electronic maintenance field presented previously.

Since troubleshooting tasks are considered the most difficult responsibility of the maintenance technician, a great amount of the research and development effort concerning electronic training has been devoted to these tasks. Simulators, such as the Tab Test (Crowder et al., 1954), the MASTS test (Grings et

Table 1. **Correlations Between Job-Task Performance Tests and Theory Tests, Job Knowledge Tests, and School Marks**

Researchers	Type of Job Task Performance Test (JTPT)	Theory Tests	Job Knowledge Tests	School Marks
Anderson (1962a)	Test Equipment JTPT			.18-.33
Evans and Smith (1953)	Troubleshooting JTPT	.24 & .36	.12 & .10	.35
Mackie et al., (1953)	Troubleshooting JTPT	.38		.39
Saupe (1955)	Troubleshooting JTPT		.55	.56
Brown et al., (1959)	Troubleshooting JTPT	.40		
	Test Equipment JTPT		.29	
	Alignment JTPT		.28	
	Repair Skills JTPT		.19	
Williams and Whitmore (1957)	Troubleshooting JTPT			
	(Inexperienced Subjects)	.23		
	(Experienced Subjects)	.15		
	Adjustment JTPT			
	(Inexperienced Subjects)	.02		
	(Experienced Subjects)	.21		
	Acquisition Radar JTPT			
	(Inexperienced Subjects)	.03	.36	
	(Experienced Subjects)	.14	.22	
	Target Tracking Radar JTPT			
	(Inexperienced Subjects)	.24	.33	
	(Experienced Subjects)	.20	.38	
	Missile Tracking Radar JTPT			
	(Inexperienced Subjects)	.09	.15	
	(Experienced Subjects)	.19	.32	
	Computer JTPT			
	(Inexperienced Subjects)	.08	.24	
	(Experienced Subjects)	.06	.14	
	Total JTPT			
	(Inexperienced Subjects)	.14		
	(Experienced Subjects)	.20		
Crowder et al., (1954)	Troubleshooting JTPT	.11	.18-.32	

al., 1953), and the similar AUTOMASTS test (Bryan et al., 1959) have been developed. (The AUTOMASTS test was essentially a refined and automatically-recording form of the MASTS test.) The Tab Test displayed a schematic of the equipment. The subject lifted a tab at each test point to obtain information normally obtained by using test equipment on the real equipment. The MASTS test used corks instead of tabs. Two studies were made concerning the effectiveness of the Tab Test. The most extensive study by Crowder et al., (1954) found correlations of .12 and .16 for two forms of the Tab Test with JTPT Evans and Smith (1953) found a correlation of .07 using similar tests. The MASTS and AUTOMAST tests were not compared with real equipment JTPT tests. Several more recent studies have indicated that the electronic technicians tested could not obtain information from their test equipments accurately (Anderson, 1962a; Foley, 19[illegible]. Perhaps, the various electronic technicians used as subjects for the above correlation studies had the same weakness. This would account in part for such low correlations.

Peer ratings and supervisors' ratings also are poor indices of technicians' capabilities to perform job tasks. Crowder et al., (1954) found that results of JTPT correlated only .16 with supervisors' ratings and .23 with peer ratings. Anderson (1962a) found various correlations from .19 to .41 between JTPT and supervisors' ratings. Wilson has indicated that:

> High intercorrelation of rating scales despite great and extensive efforts to define rating variables in mutually exclusive manners have rendered ratings disappointing to anyone doing research in the field of selection and have almost eliminated their value in training studies (Wilson, 1962, pp. 370-371).

Thorndike gives an example in which airplane commanders were rated while going through operational combat training. A rating on "likeableness" correlated highest of any of the ten traits rated with overall rating of suitability for combat flying (Thorndike, 1947). The Finucane (1966) study cited earlier indicated an insignificant correlation of .14 between JTPT of engine maintenance jobs and the commander's evaluation report, which was a rating scale. Rating scales based on supervisors' opinion of how their technicians perform have not been very successful as measuring devices.

This writer's opinion is that much more work can be done to improve symbolic maintenance tests as substitutes for JTPT. In the opinion of the writer, higher correlations could possibly be obtained by a different approach to the development of symbolic substitute tests. A study of the Tab Tests report of Crowder et al., (1954) indicates that the JTPT used as the criterion measure contained many distractions and interruptions to the subject's troubleshooting strategy (cognitive process) such as using test equipment to obtain test point information. In addition to such interruptions to the cognitive process, the subject can obtain faulty test point information by improper use of his test equipment. In the symbolic substitute Tab Tests all of these potential pitfalls of the actual task were avoided. The subject was given a printed test point readout. The injection of equivalent pitfalls into symbolic substitutes would possibly increase their empirical validity.

Such tests, however, in their current stage of development, cannot be used as substitutes for actual JTPT with great confidence. Although job knowledge paper and pencil tests about the various job tasks appear to be somewhat better than abstract theory tests for measuring subjects' ability to perform job tasks, their correlations with criterion JTPT have not been very high. The measuring devices that are being used as substitutes are certainly not good enough to replace JTPT in electronic training courses and in field maintenance units.

VI. TECHNOLOGY OF TRAINING DEVELOPMENT AND THE CONTENT OF FORMAL ELECTRONIC MAINTENANCE PROGRAMS

Modern technology for development of training programs is based on a systems approach (Eckstrand, 1964). The application of the systems approach as described by Eckstrand would result in a job oriented training program that would produce men capable of performing effectively the tasks of the jobs for which they are being trained.

A key element of the technology is the care used in deriving training objectives based on a *systematic identification* of the tasks of the job or jobs for which a training program is being developed, followed by a *systematic analysis of the identified tasks.* The second key element is as important as the first. At the time the objectives are being formulated, the criterion measures are also being developed. Based on the analysis of each identified task, the content of the training program is developed. The criterion measures are based on the *objectives* of the training program not on its *content.*

After students complete the training program, they must pass the criterion measures. A student should not graduate until he achieves the criterion. If a large number of students fail the criterion tests, the content of the training *should be* modified until a major part of the student population can achieve the criterion. In the case of the training of electronic technicians, a job task identification would result in some mix of the tasks mentioned previously as they apply to the specific hardware to be maintained. The objectives of electronic maintenance training programs, which are developed using this systems approach, should be the training of personnel to perform the tasks of their maintenance jobs effectively. Accordingly, the criterion measures should be designed to ascertain that the graduates can in fact perform these job tasks. The name now generally applied to such criterion measures which prescribe a "go, no-go" performance standard is criterion referenced measures (Glaser & Klaus, 1962).

The development of the core of current electronic maintenance training programs predates by many years the technology of training described above. Their current pattern of content is a product of the World War II maintenance training efforts (Foley, 1967b) although some of the content predates World War II. All of these courses contained the study of electronic theory followed by the study of equipment to be maintained. The equipment portion of such courses in all cases included a study of the theory of operation including circuit analysis. The equipment portion usually included "hands on" training, such as checkout, alignment and troubleshooting. The amount of such "hands on" activity varied from training center to training center and at first even from instructor to instructor.

The source of electronic theory and equipment theory of operation portion of these courses was the electronic engineering design community. These materials were academic in nature and traditional academic testing procedures were used to measure them – mostly of the paper and pencil, multiple-choice objective type. The "hands on" training content was not determined in any systematic way as would now be required by the systems approach to training development. It was usually determined by what instructors of a training program decided it should be. The measurement of each student's success was forced into the grading system used in the academic portion of the course.

During World War II, checkrooms were established in technical schools of the Army Air Force. Both written and practical (JTPT) were carefully developed and administered. In most cases the proficiency measures used were of the norm referenced type (Glaser & Klaus, 1962; Eckstrand, 1964). This pattern of training and measurement continued with improvements in job relevance until the mid-1950s.

In the late 1950s, many checkrooms were eliminated and the "hands on" equipment training was greatly reduced because of the expense. The result was modification of equipment training to the orientation, "hands off" type. The measurement of the success in such training could only be of the paper and pencil variety. At this time the electronic training programs lost practically all of their job orientation and the use of JTPT almost disappeared. The learning to perform actual job tasks was delayed until a graduate was assigned to a field maintenance unit.

The theory content of the training programs was retained with little change and each student's success in learning theory continued to be measured primarily by norm referenced paper and pencil tests *which are the appropriate instruments for measuring this type of material.* The question is how much does such learning contribute to a technician's ability to perform the tasks of his job. And such relevance should not be assumed as true without proof. (Formal mechanical maintenance training has been more job oriented than electronic maintenance training. The mechanical training retained its job orientation even when the use of JTPT was de-emphasized.)

Although action was taken beginning in 1968 to increase the use of JTPT in conjunction with equipment training, the electronic training programs have retained their great theory orientation. All of a student's scores from JTPT and from paper and pencil theory and equipment knowledge tests are combined into a single course score. Most of the tests from which the student scores are obtained are based on the *content of the training program and not from task content of the jobs* to which graduates are assigned. As a result, these training programs do not truly reflect the modern technology for course development, described previously, nor do they reflect the recommended procedures for measuring course objectives by criterion referenced tests.

VII. WHY FORMAL JOB TASK PERFORMANCE TESTS ARE NOT USED MORE EXTENSIVELY

Even though paper and pencil tests in their current state of development have been shown to be very poor substitutes for actual JTPT, the use of paper and pencil tests still persists. The question, therefore, is raised as to why JTPT are not used. Harris and Mackie (1962) report a study concerning factors influencing the use of JTPT in the Navy. They have summarized the reasons for not using such tests as follows:

> The discrepancy between the generally favorable attitudes of both supervisors and instructors toward the concept of practical performance tests and the most limited use being made of them was a result of problems connected with their development and implementation. Their infeasibility was the primary reason mentioned by supervisors and instructors for not using performance tests. A related reason – tests had not been developed or were not available – was mentioned frequently by a supervisor about operating activities.
>
> The most prominent barrier, in the minds of supervisors and instructors, to the use of practical performance tests was the degree to which the time necessary for testing must be taken from that required for carrying out the primary mission of the activity or school. Performance tests were generally felt to require too much equipment time and too much personnel time to be feasible (Harris & Mackie, 1962, pp. 4-5).

Often the amount of time required to administer a good performance test is rather great. The performance tests used in the study by Williams and Whitmore (1959), required approximately eight hours to administer. The four performance tests used in project REPAIR required a total of approximately ten hours (Brown et al., 1959). But in view of the relevance of JTPT to the job objectives of electronic training courses, the low validity of symbolic substitutes for these tests, and the high cost of *electronics* technical training, the rationale of "too little time available" is very weak.

Another serious impediment to the use of JTPT is the lack of information on how to construct good JTPT. There are many textbooks on written tests. The only text material of any magnitude that the writer has been able to find was a chapter in Micheels and Karnes (1950). The vocational education literature indicates that considerable emphasis has been placed on the development of job oriented course objectives and job relevant curricula materials. But very little emphasis or attention has been given to appropriate measurement procedures for ascertaining that these job objectives have been achieved. Instead of developing their own measurement literature, vocational educators have attempted to follow the measurement practices that have been developed by and for the academic community. This is a very serious weakness in the field of vocational education since what is formally tested is usually taught better and learned better than those curriculum items which are not tested (See Figure 1).

The Armament Systems Personnel Research Laboratory (ASPRL), at Lowry Air Force Base, Colorado, a laboratory of the Air Force Personnel and Training Research Center (AFPTRC), supported a rather comprehensive research and development effort concerning the measurement of job performance. One of the results of their work was *A Guide for Use in Performance Testing in Air Force Technical Schools* (Highland, 1955). But it was published at about the same time that checkrooms were being abolished and performance measurement was being de-emphasized in Air Training Command. Although its use in its present form is not recommended, there is certainly a need for such a guide on how to develop and use criterion referenced JTPT. As noted by Eckstrand (1964), this is a soft area in the modern technology of training. It is too much to expect technical training personnel to prepare good JTPT without adequate in-depth guidance. (AFPTRC and ASPRL were abolished in 1958. As a result, this important measurement work was discontinued. For a short summary of the work of AFPTRC see Glaser, 1964.)

VIII. THE JOB VALIDITY OF THEORY IN ELECTRONIC TECHNICIAN COURSES

It is often contended that a knowledge of theory is an essential part of an electronics technician's training since such knowledge gives the technicians the broad background necessary to perform many types of maintenance activities. The published literature presents no hard data to support this widely held contention. However, there are several published studies which support the hypothesis that a knowledge of theory as measured by paper and pencil tests does not insure the ability to perform maintenance tasks. In fact, there is some convincing evidence that there is little relationship between true ability to perform job tasks and knowledge of theory as measured by paper and pencil tests.

Lyman (1971) makes the distinction between maximum performance tests and typical performance tests. Performance tests as used in Lyman's discussion have reference to all types of performance whether they be verbal, manipulative, or complex. Since the word "performance" is used in conjunction with job task performance tests, the term "effort" is used here in order to avoid confusion. When a formal examination is given by a teacher, each student usually makes a rather intense effort to prepare for the examination and also makes a maximum effort during the examination. Micheels and Karnes (1950) make interesting comments about this effect of tests on learning.

> *Tests Provide an Incentive for Application*: In order to gain an understanding of the interest of students in test results, one has only to observe their reactions when test papers are returned for review and discussion. If there is any question about this point, consider your own experiences when you were on the receiving end of the tests. Think of the times you have put forth extra effort in getting ready for a test. Some readers will still be very much concerned with their ability to pass the tests that are constructed by their instructors. They will agree with the authors that knowledge of a forthcoming test is a powerful motive to start studying.
>
> It would be nice (perhaps) if all students in a school were interested in learning all they possibly could whether or not a check were made on their progress. This, however, is not the case. A few will put forth their best efforts whether or not tests are given. But the majority will work harder if they know that they are to be held accountable for what has been taught. Generally, the instructor who administers the most rigid programs of evaluation gets the greatest amount of work out of his students.
>
> There is one danger in using tests and test results as an incentive to students to apply themselves to work and study. Their interest in marks can be a superficial one which easily leads to efforts to "hit the test" rather than learn the subject matter for its value now and in the future. Students who study primarily to pass tests usually forget the material much faster than those who are interested in learning because of the values to be derived. A positive suggestion is this: Give rigid tests; give them frequently; but design tests that require your students to make application of what has been taught (Micheels & Karnes, 1950, p. 89).

Students tend to study those things most likely to appear on examinations. Therefore, it follows that students in *electronic* courses usually study theory and job knowledge, rather than practice job performance. All such courses use paper and pencil tests of theory and job knowledge. It is safe to assume that at least some electronic technology courses can be found in which no formal examinations or tests are given concerning the students' ability to perform job tasks.

In other courses each student may perform the same tasks in his laboratory or shop practice, tasks, which should appear in job task performance tests. In lieu of a performance test score, he will be given a grade or score on the work of his laboratory manual. Ordinarily, the work done in the laboratory is of the "typical" effort type rather than the "maximum" type (Lyman, 1971). The student is not faced with the problem of having to demonstrate his ability to perform these tasks in a formal examination. Under these conditions, he will not spend as large an amount of time vigorously practicing job behaviors as he will practicing the verbal behaviors required by the formal written examinations. The student also makes a special effort during the JTPT. The writer has observed students practicing such task behaviors on their own time in preparation for such performance tests. Teachers also tend to place more emphasis on teaching formally tests behaviors than they do on teaching those behaviors evaluated in some less demanding manner.

Figure 1. **Effect of tests on learning effort.**

Williams and Whitmore (1959) reported a study in which a verbal theory examination and a job sample performance test were given to graduates of an Army electronic maintenance course. A written theory and job knowledge test was developed to measure the retention of the knowledge acquired by technicians during their school training in electronics. The performance test was made up of actual job problems using actual field equipment. Both tests were administered to 91 graduates at the time they completed school, and both tests were administered to another group of 98 graduates with from 1 to 57 months of field experience (mean 19 months) who had previously completed the same training. Among the resulting conclusions were the following:

> a. The field-experienced group scored higher on the performance test but lower on the written test than did the inexperienced group, with a trend toward higher performance and lower written scores as the amount of field experience increased, and b, the contrasting results in the written tests scores following graduation raise a question as to whether some of the material in the NIKE AJAX IFC Maintenance Course is relevant to the field job (Williams & Whitmore, 1959, p. iii).

The written examination used by Williams and Whitmore was based on the content of the United States Army NIKE course and the performance test used sampled the actual job behaviors. The correlation between the results of paper and pencil theory test and JTPT are very low. Williams and Whitmore found only a .144 correlation between basic electronic (theory) paper and pencil test scores and JTPT scores for inexperienced graduates. For experienced graduates, the correlation was .196. A student's performance on tests of knowledge of theory seems to tell very little about his ability to perform job tasks. Also, it does not predict his ability to learn such job tasks very well. The first of five recommendations of this study was that:

> The Air Defense School review the training objectives for the NIKE AJAX IFC technicians course and analyze current course content to determine its job relevance, and some form of the performance tests developed in this study be used to evaluate the effects of any changes made in the NIKE AJAX course (Williams & Whitmore, 1959, p. iii).

Several other studies, such as Brown et al., (1959), and Shriver (1960), cast further light on the lack of job relevance in much of the theory content of electronic technology courses. Anderson (1962b) made a study of the amount of mathematics used by Navy electronic technicians and found that very little was actually used on the job. The mathematical content of these courses makes them very difficult for many students.

Examples of lack of relevance in many curriculum items of the traditional theory courses can be demonstrated by careful analyses of these items. A casual, uncritical examination of the same items could result in the conclusion that they are relevant. The subject of the vacuum tube voltmeter is presented as an example. On the job, the technician is required to use this meter daily, and he should be able to use it precisely. In a job oriented training program, the student learns to use a vacuum tube voltmeter by practicing its various uses. He learns to hook up the meter to the proper points in the equipment being tested, to select the proper voltage range of the meters switch, and to read voltage values accurately on the proper scales of the meter display. In this situation, he pays little attention to the internal theory of operation of the vacuum tube voltmeter, much as the driver of an automobile normally pays little attention to how the engine in his automobile works while he is driving.

In a theory course, under the same heading of vacuum tube voltmeter, the student learns a different set of behaviors — the ability to verbalize and to analyze how the circuits of the meter are designed using a symbolic representation or schematic. It is possible for him to learn such behaviors and never actually use or see a voltmeter. The latter behaviors are seldom, if ever, required of the technician on his job. Although such verbal knowledge probably does the student no harm, it certainly contributes little to his ability to use the vacuum tube voltmeter in job situations. Foley (1969) reports that Air Force technicians, who had been upgraded on the basis of a job knowledge tests, could not use test equipment accurately when given a JTPT. Anderson (1962a) reports a similar situation in the Navy.

Finucane (1966) reports an Army situation of an apparent spare parts shortage for the repair of engines. An investigation indicated that repairmen were discarding good parts as being faulty. The investigation also indicated by use of JTPT that the repairmen could not locate and identify defective parts even though they had been upgraded on the basis of a job knowledge test. Elliott and Joyce (1968) and Shriver, Fink, and Trexler (1964) have demonstrated that electronic job tasks can be performed by subjects with little or no knowledge of traditional theory or principles. Such subjects, however, must be proficient in the use of test equipment, handtools, and well designed job performance aids.

The job oriented studies summarized in Foley (1967b) indicated that students graduating from job oriented courses, in which they received considerable practice in performing job tasks, were able to perform the assigned job tasks almost immediately. Students of similar aptitude, graduating from theory oriented courses were not able to perform the same job tasks until after five or six months of experience on the job. The results of a number of independent critical examinations of the theory content of various Armed Forces electronic technician courses, in light of actual job content, are the bases for concluding that traditional theory content has doubtful relevance to the tasks of jobs for which personnel are being prepared.

In view of existing evidence, the validity of the contention that a knowledge of electronic theory is a necessary prerequisite for the performance of electronic maintenance tasks is suspect. Even though this contention is deeply imbedded in the culture of the electronic maintenance community (Figure 2), it certainly should be subjected to a large scale, objective, controlled evaluation to determine its empirical validity. However, as long as the learning of electronic theory is a stated objective of electronic maintenance courses, students should be tested as to how well they learn this subject matter. The appropriate instruments for such measurement are paper and pencil tests. *But the results obtained from such paper and pencil measurements should be reported separately from the results obtained from criterion referenced JTPT.*

IX. GRADES, MARKS, AND TEST SCORES

The American school tradition requires some form of a grading system. A search of the literature in the *Education Index* reveals that a great many articles are produced each year on this subject. No references were found in the same index regarding the criterion problem, the answers to which are basic to good testing and grading procedures. Stanley (1964) says that "changing scores into grades is at best a rather arbitrary process, and this arbitrariness is further complicated by the public-relations aspects of reporting to parents." Grades, however, can be no better than the test procedures that produce them and because of their limitations, probably not as good. Even if the tests used in a course do reflect the objectives of the course, grades will not be completely satisfactory as Travers (1950) indicates: "Any system of assigning grades inevitably includes many unsatisfactory elements since the scale does not, and possibly cannot, meet the criteria of a satisfactory measuring device."

Stanley (1964) indicates that most modern report cards contain both grades and checklist items. Scholastic achievement is usually measured by a letter system such as "A, B, C, D, F," while citizenship is indicated by an adjective system such as S (Superior), N (Normal), and U (Unsatisfactory). He indicates that achievement, as measured by tests, should not be combined with characteristics evaluated subjectively such as "effort," punctuality, deportment, and neatness of work. If the results of testing are to be reported accurately in the grading system used for electronic training courses, the achievement grades should not reflect effort. Such a practice decreases the validity of the grades as indicators of a student's proficiency in the job task performance objectives of the course.

As previously discussed, criterion referenced JTPT (used to measure ability to perform electronic maintenance job tasks) and norm referenced paper-and-pencil tests (used to measure electronic theory and job knowledge) are to a great extent measuring different factors and are based on different measurement technology. Grades derived from each source should be reported separately. This would permit the users of grades to make separate judgements about these two aspects of the training program as indicated by Frederiksen.

> Criterion measures which more accurately reflect the objectives of instruction and which permit judgements to be made separately about various aspects of the teaching program are needed.
>
> In recent years there seems to have been a rapidly growing awareness of the need for training evaluation and for developing proficiency measures for use in evaluation which better reflects the sought-for outcomes of instruction (Frederiksen, 1962; p. 323).

Unless separate grades are reported, it will be impossible to determine the effectiveness of the theory aspects and the job aspects for electronic technology education in job success follow-up studies. Separate grades also can be justified on the bases of better communications. Travers (1950) indicates that:

Due to traditional influences the elimination of traditional theory is probably impossible at this time. Perhaps substantial reductions could be made gradually. Some of the real reasons for the persistent demands for theory should be examined. In spite of the experimental evidence to the contrary, many electronics teachers and many employers of graduates demand a knowledge of traditional electronic fundamentals. Pickering and Anderson found this attitude among supervisors in a 1966 follow-up study of graduates who had completed an experimental performance-oriented electronic-technology course in the Navy.

> Preliminary follow-up results indicate that the experimental ET (electronic technicians) graduates have been generally well received in the fleet. However, it is also clear that a subtle but widespread bias against them exists. For example, several supervisors have made comments similar to this: "X is a good man. He's done everything I've asked him to do. Of course, I wouldn't assign him to a really difficult job because he hasn't had enough theory." Judgments of limited capabilities are being made independent of objective evidence. Regular ET A school graduates have successfully completed a complex and difficult training program that has placed emphasis on the attainment of knowledge about theoretical concepts and demonstration of capacity to manipulate electronic relationships mathematically. ET's as a group, are justly proud of their accomplishments. The experimental ET school graduates are not considered part of this group. What effect these attitudes will have on the careers of the experimental ET's is not yet evident, but the attitudes are real enough (Pickering & Anderson, 1966, p. 39).

If the supervisor is an electronic technician, he has successfully completed a rather lengthy and difficult electronic theory or fundamentals course and he now is able to perform the job tasks of an electronic technician, at least sufficiently well to "get by." Even though he probably learned to perform most of these tasks on the job over a period of several years, he may firmly believe that completing the difficult theory course is the key to his success as an electronic technician. A new man is not one of the "in group" unless he has been initiated in the same way. The new man will not be permitted to demonstrate his ability to perform until he demonstrates that he can give "lip service" to the theory. The writer has experienced this same unquestioned belief in the value of theory in his contacts with large numbers of Air Force, Army, and Navy school and field personnel. He also has found the same bias among most of the civilian electronic technology course instructors he has contacted. The requirement that technicians have such a knowledge of theory has been further reinforced in the Department of Defense by the theory-oriented job knowledge tests that each person must pass before he is promoted from apprentice to specialist to technician.

Figure 2. **Why theory training?**

A grade system is a measuring scale, but before it can be used as such, it is necessary to know what it is measuring. At the present time, student grades are used to measure at least three different things. The grades given by one teacher may indicate the extent to which the goals of the course have been achieved in the students. The grades given by a second teacher may indicate the relative amount of progress made by a student. The grades given by a third may indicate the relative standing of the students (grading on the curve). According to the system of the second teacher, a student who starts a course with a poor background and finishes with average proficiency may deserve a better grade than one who starts out with a good background and finishes the course with a high level of achievement. A school should determine what is to be measured by grades so that those who use the cumulative records may know how to interpret them. Most cumulate records cannot be interpreted because the grades of different teachers mean different things (Travers, 1950, p. 145).

Scores for electronic job task performance tests have long been a problem especially where discrimination is desired. In course achievement tests, "go, no-go" standard of performance within a time limit can be applied, as reported by Williams and Whitmore (1959). They obtained a differential score by converting time required to successfully complete the task into a standard score. Hansen (1958) who has studied the relative merits of electronic troubleshooting job-task performance test scores, has recommended that several scores be used:

1. "Time required to complete the checklist activities in symptom-seeking phase of the test." The writer has classified this type of behavior as checkout procedures.

2. "Time required to complete the trouble-location phase of activity." This is similar to the Williams and Whitmore (1959) procedure.

3. "Number of discrete steps of trouble-locating activity."

4. "Number of erroneous replacement units." (Hansen, 1958, pp. 41-42).

None of these suggestions is truly satisfactory. *The whole problem of how to score and report the results of JTPT is certainly a worthy area of exploratory development.* One promising solution to the discrimination problem has been suggested by Dr. Gordon Eckstrand of the Advanced Systems Division of the Air Force Human Resources Laboratory, i.e., the *training time* that is required by a student to reach a "go, no-go" standard of performance for each task tested. A variation of this approach would be the number of testing attempts required to reach criterion on each JTPT. In the meantime, whatever scores are used to determine the theory grades and job task performance grades, and the meaning of these two types of grades based on different scoring systems should be adequately communicated to the users of these grades. In order to further increase the meaningfulness of the job task performance grade, the writer would recommend that it be supplemented by a list of activities included in the course job task performance measurement system, together with an adjective rating of each student's quality of performance for each activity. Such a profile of a student's job task proficiency would also emphasize the importance of this course objective to the students, instructor, school administrators, and maintenance supervisors.

X. SUMMARY AND CONCLUSIONS

1. This report represents an in-depth analysis of the state-of-the-art concerning measurement procedures for ascertaining the job proficiency of electronic maintenance personnel. It is based primarily on a review of the literature made as part of the writer's doctoral dissertation (1967a). It also reflects several other efforts and events supported by the Advanced Systems Division, Air Force Human Resources Laboratory, Wright-Patterson Air Force Base, Ohio, including a bibliography on the measurement of maintenance personnel prepared by Dr. Askren (1963), a paper on the status of technology of training by Dr. Eckstrand (1964), a field survey of maintenance practice (Folley & Elliott, 1967), a 1966 conference on the *Assessment of Complexed Operational Tasks* arranged and reported by Dr. Chiles (1967), and a 1968 conference on human factors testing arranged and chaired by Mr. Snyder (Snyder et al., 1969).

2. The first consideration that should be addressed when a test is being developed is what is it supposed to measure. Whatever it is supposed to be measuring is its criterion. In the case of tests concerning the ability to perform electronic maintenance, the ultimate criterion, therefore, is the tasks of electronic maintenance performed in their job environment. For testing purposes it is usually impossible to simulate all of the conditions of such an ultimate criterion and at the same standardize test conditions. But it is possible to

require the test subject to demonstrate that he can perform key tasks of this job removed from their actual job environment. A performance test battery made up of such tasks is considered to be very close to the ultimate criterion. The difference between the real world and the simulation of the real world for testing purposes is called the criterion problem. Many times paper and pencil job knowledge tests or supervisors' ratings are substituted for near ultimate criterion tests, making the criterion problem greater. The farther removed from the criterion that testing procedures become, the more the danger that the tests are not truly measuring what they are supposed to be measuring. The degree to which tests relate to the criterion is called the degree of empirical validity.

3. This report indicates that two very serious criterion problems exist in many electronic training programs. The empirical validity of the paper and pencil tests used to measure student achievement has not been determined and the job relevance of much of the training content has not been ascertained.

4. The full application of modern training development technology based on a systems approach supplies a realistic solution to both the problems. A key element of the technology is the care used in deriving training objectives based on a systematic identification of the tasks of the job or jobs for which the training program is being developed, followed by a systematic analysis of each identified task. A second key element is as important as the first. At the time the objectives are being formulated the criterion measures are also being developed. Based on the analysis of each identified task, the content of the job oriented training program is developed. The criterion measures are based on the *objectives* of the training program, not on the *content* of the training program. The criterion measures, appropriate for determining student achievement in electronic maintenance courses developed in this manner, are Job Task Performance Tests (JTPT).

5. The development of the core content and structure of current electronic maintenance training programs predated by many years the modern course development technology. The core content of these courses reflect the influence of the electronic design engineering community. The courses are theory oriented and reflect the subject matter structure found in most electronic engineering programs. The tests used to determine student achievement are based on the content of this academic subject matter and are of a paper and pencil variety, which of course are appropriate for measuring academic achievement.

6. In addition to this more or less standard theory content, current electronic maintenance training also contains instruction on typical electronic equipments or sets. Such instruction may be theory oriented or job oriented or both. But the job orientation of such training programs has not been achieved in the careful manner recommended by the modern technology for training development. Some JTPT are administered but the main thrust of measurement is of the paper and pencil job knowledge type. The scores obtained from all theory, job knowledge, and performance tests are combined to produce a single score indicating school success. This is an inappropriate action. Most paper and pencil tests measure student success in learning theory and job knowledge, and are norm-referenced. JTPT measure ability to perform key job tasks and are criterion-referenced. Mixing the two types of scores gives a meaningless result.

7. The graduate of such training programs is assigned as an apprentice to a field maintenance unit where he receives on-the-job-training which is certainly job oriented. However, he seldom if ever is given formal JTPT to determine how well he can perform the key tasks of his job assignment. His advancement to the journeyman level is greatly dependent on his ability to pass a paper and pencil Specialty Knowledge Test which is of the job knowledge test variety.

8. The above analysis of current electronic training and testing practice indicates that great reliance is placed on paper and pencil theory and job knowledge tests for determining training and job success almost to the exclusion of criterion JTPT. This volume identifies, consolidates and reviews the data found in the available reported research concerning the empirical validity of paper and pencil theory and job knowledge tests and school marks. The correlations, obtained when the results of most of these tests have been compared with the results of applicable JTPT, are very low. This would indicate that the empirical validity for most of these paper and pencil instruments as well as school marks obtained from combining their results is suspect. The limited amount of research literature found concerning peer ratings and supervisor's ratings also would indicate that their empirical validity is suspect.

9. Several paper and pencil tests concerning troubleshooting such as the Tab Test and the Multiple-Alternative Symbolic Trouble (MAST) Test are reported in the literature. It would seem that such tests would test the cognitive factors of troubleshooting. However, the comparison of the results of these tests and the results obtained from JTPT reflecting the same equipment troubles produced rather low

correlations. In the opinion of the writer, the JTPT contained many distractions, normally present on the job, that the tests based on the pure cognitive processes did not contain. Considering this, the empirical validity of such paper and pencil symbolic substitutes could possibly be improved by further development and refinement. In view of the expensive equipment requirements for most JTPT, the development of symbolic substitutes of high empirical validity would be very desirable.

10. There are many studies concerning the relative effectiveness of theory oriented and job oriented electronic maintenance training reported in the literature. The results of these studies indicate that the graduates of the job oriented training programs are able to perform productive work immediately upon assignment to field maintenance units. The criterion of school success in these courses is the ability to pass JTPT, rather than paper and pencil tests on theory and job knowledge. The percent of students of high electronic aptitude who successfully complete either theory oriented or job oriented training is about the same. A much higher percent of students of medium (not low) aptitude successfully complete job oriented training. In some cases almost no medium aptitude students are able to successfully complete the theory portion of the theory oriented programs.

11. Such results would indicate that much of the theory content of traditional electronic training programs is not relevant to the performance of the maintenance tasks performed by personnel in their first enlistments. However, the contention that a knowledge of theory is a necessary prerequisite to the successful performance of maintenance is deeply imbedded in the culture of the electronic maintenance community. But there is no doubt that this prerequisite adds greatly to the personnel costs of the Air Force.

12. This report has identified many problems and questions concerning Air Force maintenance. These problems and questions concern the adequacy and efficiency of the personnel selection procedures, the personnel training procedures, and field maintenance practices. Since most of the reported research has concerned electronic maintenance, this report has been concerned primarily with this type of maintenance. However, much of this research would also apply to mechanical maintenance.

13. The findings and analyses of this volume indicate an urgent requirement for well planned, adequately funded, comprehensive exploratory and advanced development programs for systematically addressing and solving these problems. The first effort of such a program would be a comprehensive and in-depth study, based on hard data, concerning how well maintenance men can perform the key tasks of their jobs. Such an in-depth study would provide a needed base and feedback for follow on studies for the improvements of maintenance practice, personnel selection, and personnel training.

14. A serious gap in the modern technology of training development, is the lack of adequate guidance on how to develop and administer criterion referenced JTPT. Prior to the work on criterion referenced JTPT, reported in Volume II of this report, very little exploratory development on JTPT technology had been supported since the late 1950s. There is an urgent requirement for a modern guide on how to develop JTPT for military technical training and civilian vocational training.

XI. ACTIONS AND RECOMMENDATIONS

Actions

Based on the analyses and conclusions presented in this volume, several actions have already been taken.

1. A contractural effort resulted in the development of Job Task Performance Tests (JTPT) and appropriate scoring schemes for each of the key electronic maintenance job activities identified in this volume (Contract F33615-69-C-1232). The work statement for this contract is supplied as Appendix A.

2. A follow-on contractual effort resulted in the development of a detailed Test Administrators Manual, a training program for Test Administrators and a limited tryout of the JTPT (Contract F33615-70-C-1695). The results of both these JTPT contractual efforts are reported in AFHRL-TR-74-57 (II), (Parts I and II).

3. Two other follow-on contractual efforts were initiated, which resulted in the development and limited validation of both graphic and video symbolic substitutes for each type of job activity for which JTPT had been developed (Contracts F33615-70-C-1550 and F33615-71-C-1505). The results of these efforts are reported in AFHRL-TR-74-57 (Volumes III and IV).

Recommendations

1. Comprehensive exploratory and advanced development programs should be established and adequately funded that systematically identify and solve problems concerning maintenance practice in the field, and problems concerning the selection and training of maintenance personnel.

2. A necessary first effort of such programs should be a comprehensive and in-depth study, based on hard data, concerning how well maintenance men can perform the key tasks of their jobs. Such an in-depth study would require the development and use of JTPT for electronic and mechanical maintenance jobs. The results would produce criteria for determining the validity of personnel selection and training programs.

3. A modern guide on how to develop JTPT for maintenance jobs should be developed. Such a guide should be an on-going effort. This guide should be developed now, based on our current "know how." The guide should be updated periodically as more expertise and sophistication in the development and use of such tests are achieved.

4. In training programs in which both paper and pencil tests and JTPT are administered, their results should be reported separately. The results of JTPT should be reported in the form of a profile rather than under the umbrella of a single meaningless score.

REFERENCES

Anderson, A.V. *Training, utilization and proficiency of Navy electronics technicians: III. Proficiency in the use of test equipment.* Navy Technical Bulletin 62-14, AD-294 330. San Diego: U.S. Navy Personnel Research Activity, 1962. (a)

Anderson, A.V. *Training, utilization and proficiency of Navy electronics technicians: IV. Proficiency in mathematics.* Navy Technical Bulletin 62-16, AD-296 541. San Diego: U.S. Navy Personnel Research Activity, 1962. (b)

Askren, W.B., Jr. *Bibliography on maintenance personnel performance measurement.* AMRL Memorandum Report P-45, AD-439 610. Aerospace Medical Research Laboratories, Wright-Patterson AFB, Ohio, June 1963.

Brown, G.H., Zaynor, W.C., Bernstein, A.H., & Shoemaker, H.A. *Development and evaluation of an improved field radio repair course.* Technical Report 58, Project Repair. Human Resources Research Office, Washington: The George Washington University, 1959.

Bryan, G.L. The training of electronics maintenance technicians. In R. Glaser (Ed.), *Training Research and education.* Pittsburgh, Pa.: University of Pittsburgh Press, 1962, 303-09.

Bryan, G.L., Bond, N.A., Jr., & La Porte, H.R., Jr. *An experimental battery for measurement of the proficiency of electronics technicians.* Technical Report No. 12, Department of Psychology, University of Southern California, March 1955.

Chiles, W.D. *Assessment of complex operator performance.* AMRL-TR-67-239, AD-681 539. Aerospace Medical Research Laboratories, Wright-Patterson AFB, Ohio, August 1967.

Crowder, N., Morrison, E. J., & Demaree, R. G. *Proficiency of Q-24 radar mechanics: VI. Analysis of intercorrelations of measures.* AFPTRC-TR-54-127, AD-62 115. Lackland AFB, Tex.: Air Force Personnel and Training Research Center, 1954.

Eckstrand, G.A. *Current status of the technology of training.* AMRL-TR-64-86, AD-608 216. Aerospace Medical Research Laboratories, Wright-Patterson AFB, Ohio, September 1964.

Elliott, T.K., & Joyce, R.P. *An experimental comparison of procedural and conventional electronic troubleshooting.* AFHRL-TR-68-1, AD-681 510. Wright-Patterson AFB, Ohio: Training Research Division, Air Force Human Resources Laboratory, November 1968.

Evans, R.N., & Smith, L.J. *A study of performance measures of troubleshooting ability on electronic equipment.* Illinois: College of Education, University of Illinois, October 1953.

Finucane, J.L. *Measuring troubleshooting skills of maintenance personnel by paper and pencil tests.* Waukegan, Ill.: 8th Annual Military Testing Association Conference, September 1966.

Foley, J.P., Jr. *Critical evaluation of measurement practices in post-high school vocational electronic technology courses.* AD-683 929. Doctoral dissertation, University of Cincinnati, 1967. (a)

Foley, J.P., Jr. Factors to consider in developing new test and evaluation techniques. *Proceedings of the Human Factors Testing Conference, 1-2 October 1968.* Snyder, M.T. (Chm.), Kincaid, J.P., & Potempa, K.W. (Ed.), AFHRL-TR-69-6, AD-866 485. Air Force Human Resources Laboratory, Wright-Patterson AFB, Ohio, October 1969.

Foley, J.P., Jr. *Performance testing: testing for what is real.* AMRL Memorandum Report P-42, AD-412 776. Aerospace Medical Research Laboratories, Wright-Patterson AFB, Ohio, June 1963.

Foley, J.P., Jr. Summary of armed forces electronic technical education studies. Appendix K in *Critical evaluation of measurement practices in post-high school vocational electronic technology courses.* AD-683 729. Doctoral dissertation, University of Cincinnati, 1967, 192-199. (b)

Folley, J.D., Jr., & Elliott, T.K. *A field survey of electronic maintenance technical data.* AMRL-TR-67-159, AD-666 990. Aerospace Medical Research Laboratories, Wright-Patterson AFB, Ohio, November 1967.

Frederiksen, N. Proficiency Tests for Training Evaluation. In R. Glaser (Ed.), *Training research and education.* Pittsburgh, Pa.: University of Pittsburgh Press, 1962.

Glaser, R. Implications of training research for education. In E.R. Hilgard (Ed.), *Theories of learning and instruction: The sixty-third yearbook of the National Society for the study of education, Part I.* The University of Chicago Press, Chicago, Ill.: 1964, 153-181.

Glaser, R. (Ed.). *Training research and education.* Pittsburgh, Pa.: University of Pittsburgh Press, 1962.

Glaser, R., & Klaus, D.J. Proficiency measurement: Assessing human performance. In R. Gagne (Ed.), *Psychological principles in systems development.* Holt, Rinehart and Winston, New York: 1962, 419-476.

Goffard, S., Heimstra, N.W., Beecroft, R.S., & Openshaw, J.W. *Basic electronics for minimally qualified men: An experimental evaluation of a method of presentation.* Technical Report 61, Project REPAIR. Human Resources Research Office, Washington: The George Washington University, 1960.

Grings, W.W., Rigney, J.W., Bond, N.A., & Summers, S.A. *A methodological study of electronic trouble shooting skill: II. Intercomparisons of the MASTS test, a job sample test, and ten reference tests administered to fleet electronics technicians.* Navy Technical Report No. 10. Department of Psychology, University of Southern California, August 1953.

Hansen, O.K., An evaluation of some performance test scores used in electronic troubleshooting research. Unpublished Master's dissertation, Department of Psychology, University of Utah, 1958.

Harrington, G.M. *Technical education.* Vocational and technical education + reviews of the literature for six year period since issuance of Vol XX, No. 4, October 1950. Vol XXVI, No. 4. October 1956.

Harris, D. & Mackie, R.R. *Factors in Influencing the Use of Practical Performance Tests in the Navy.* Navy Technical Report No. 703-1, AD-285 842. Washington: Office of Naval Research, 1962.

Highland, R.W. *A guide for use in performance testing in Air Force technical schools.* ASPRL-TM-55-1, AD-65 480. Armament Systems Personnel Research Laboratory, Lowry AFB, Colo., January 1955.

Jenkins, J.G. Validity for what? *Journal of Consulting Psychology.* Vol X, No. 2. March-April 1946.

Lyman, H.B. *Test scores and what they mean.* Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1971.

Mackie, R.R., Wilson, C.L., & Buckner, D.N. *Practical performance test batteries for electricians mates and radiomen developed in conjunction with a manual for use in the preparation and administration of practical performance tests.* AD-98 239. Los Angeles, Calif.: Management and Marketing Research Corporation, June 1953.

Micheels, W.J., & Karnes, M.R. *Measuring education achievement.* New York: McGraw-Hill, 1950.

Pickering, E.J., & Anderson, A.V. *A performance-oriented electronics technician training program: I. Course development and implementation.* Technical Bulletin STB 67-2. California: U.S. Navy Personnel Research Activity, August 1966.

Psychological Abstracts. American Psychological Association, Inc., Washington, D.C.: Published monthly.

Saupe, J.L. *An analysis of troubleshooting behavior of radio mechanic trainees.* AFPTRC-TN-55-47, AD-99 361. Air Force Personnel and Training Research Center, Lackland AFB, Tex.: November 1955.

Shriver, E.L. *Determining training requirements for electronic system maintenance: Development and test of a new method of skill and knowledge analysis.* Technical Report 63, Project FORECAST. AD 239 416. Human Resources Research Office, Washington, D.C.: The George Washington University, 1960.

Shriver, E.L., Fink, C.D., & Trexler, R.C. *FORECAST.* Research Report 13, AD-441-248, Alexandria, Va.: Human Resources Research Office, The George Washington University, May 1964.

Snyder, M.T., Kincaid, J.P., & Potempa, K.W. *Proceedings of the human factors testing conference, 1-2 October 1968.* AFHRL-TR-69-6, AD-866 485. Air Force Human Resources Laboratory, Wright-Patterson AFB, Ohio, October 1969.

Stanley, J.C. *Measurement in today's schools.* (4th ed.) Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1964.

Super, D.E., & Crites, J.C. *Appraising vocational fitness by means of psychological tests.* (Rev. ed.) New York and Evanston: Harper and Row Publishers, 1962.

Thorndike, R.L. *Research problems and techniques.* Research Report No. 3. Army Air Force Aviation Psychology, Washington, D.C.: Government Printing Office, 1947, 50-51.

Thorndike, R.L. *Personnel selection.* New York: John Wiley and Sons, Inc., 1949.

Travers, R.M.W. *How to make achievement tests.* New York: The Odyssey Press, Inc., 1950.

Travers, R.M.W. *Educational measurement.* New York: The MacMillan Co., 1955.

Wallace, S.R. Criteria for what? *American Psychologist,* June 1965, 20(6), 411-417. (a)

Wallace, S.R. The relationship of psychological evaluation to needs of the department of defense. *Proceedings of 7th Annual Military Testing Association Conference.* San Antonio, Tex.: October 1965, 1-10. (b)

Williams, W.L., Jr., & Whitmore, P.G., Jr. *The development and use of a performance test as a basis for comparing technicians with and without field experience. The NIKE AJAX AFC maintenance technician.* Technical Report 52, AD-212 663. Human Resources Research Office, Washington, D.C.: The George Washington University, January 1959.

Wilson, C.L. On-the-job and operation criteria. In R. Glaser, (Ed.), *Training research and education.* Pittsburgh: University of Pittsburgh Press, 1962.

Wilson, H.W. Company, *Education index,* New York: Published monthly.

APPENDIX A: WORK STATEMENT FOR CONTRACT F33615-69C-1232 CONCERNING DEVELOPMENT OF CRITERION REFERENCED JOB TASK PERFORMANCE TESTS

This work statement reflects the findings and recommendations of AFHRL-TR-74-57(I) as well as other efforts and events of the Advanced Systems Division of the Air Force Human Resources Laboratory indicated in that volume. This 8 November 1968 version of the work statement was included in the contract. However, the first draft was completed in June 1968. It is included here to indicate the information and directions made available to the contractor at the time he started his exploratory development effort.

Date: 68 Nov 08

DEVELOPMENT OF JOB-TASK PERFORMANCE MEASUREMENT PROCEDURES FOR ELECTRONICS MAINTENANCE ACTIVITIES

1.0 INTRODUCTION (Objective): The primary purpose of the effort is to develop adequate procedures for measuring the ability of subjects to perform each type of job activity described and to develop an appropriate scoring scheme for use with each measure. It is anticipated that these measuring procedures and scoring schemes will have general applicability for measuring an individual's ability to perform maintenance activities for a particular weapon system or equipment. These job-task performance tests will be used in later exploratory development as criteria for future development of symbolic tests.

2.0 SCOPE: The vehicle for developing these measurement procedures will be the Radar Set AN/APN 147, the accompanying computer set AN/ASN 35, and their associated test equipment. Job task performance tests and appropriate scoring schemes shall be developed for job tasks activities associated with this electronic system, for information gathering activities (using test equipment), and for handtool activities. The test procedures used for measuring the job-activities associated with the electronic equipment should be applicable to any electronic system but the tests will be developed for the AN/APN 147-AN/ASN 35 system. The performance tests concerning the use of each test equipment will include all the capabilities of the test equipment. The handtool tests will be limited to the handtool activities associated with the AN/APN 147-AN/ASN 35 system.

3.0 GENERAL BACKGROUND: In present day electronics, technical training programs, many formal paper and pencil tests are utilized for measuring and predicting student achievement. Research in ascertaining the effectiveness of such paper and pencil predictive and achievement tests, in ascertaining the effectiveness of various training methods, and in determining on-the-job effectiveness of electronics repairmen and technicians has been hampered by a lack of job-task performance measurement and scoring procedures for various job activities. Recent study of these activities has resulted in a plan for developing effective measurement procedures. Rather than developing one test and one scoring procedure for the entire range of job activities, the development of an appropriate testing procedure and scoring scheme for each of several classes of job activities of electronic technicians is proposed. This exploratory development effort is but the first step of a plan for improving the effectiveness of Air Force measurement. The successful development of valid and workable procedures will provide the information for greatly improving the effectiveness of training and of job performance aids. They will provide a criteria for ascertaining the effectiveness of current predictive and achievement measurement practices. They will also be used as criteria for future exploratory development of valid paper and pencil substitutes.

3.1 Applicable Documents List: A comprehensive search has been made of the measurement literature including a DDC bibliography search. The most applicable technical reports, books, and articles are listed.

3.1.1 Technical Reports. These reports are on file at the Defense Documentation Center (DDC), Cameron Station, Alexandria, Virginia 22314. Each report's DDC number is included in its description below; i.e., AD ______. They can also be obtained from the Clearinghouse for Federal Scientific and Technical Information (CFSTI), Sills Building, 5285 Port Royal Road, Springfield, Virginia 22151. (Since all of the references which appeared here in this work statement are included in the references of this volume, only the author and year of publication are given here and include: Anderson (1962b), Brown et al. (1959), Crowder et al. (1954), Grings et al. (1953), Harris and Mackie (1962), Mackie et al. (1953, and William and Whitmore (1959). In addition, the contractor was given a copy of Foley (1967a).)

3.1.2 Books. (Frederiksen (1962), Super and Crites (1962), Thorndike (1949) and Wilson (1962).)

3.1.3 Articles. (Jenkins, 1946; Wallace, 1965)

4.0 TECHNICAL REQUIREMENTS/TASKS:

4.1 Areas of Consideration: It is emphasized that the primary objective of this effort is the development of adequate measuring procedures and appropriate scoring schemes for the various job activities of electronic technicians. These procedures and scoring schemes should have general applicability for measuring job effectiveness in any electronic maintenance career field. An electronic maintenance technician performs maintenance tasks on an electronics system or group of systems making use of the capabilities of electronic test equipment to obtain information about his systems. These maintenance tasks include the use of appropriate handtools. This effort is aimed at the diagnostic measure of all the activities included in the maintenance process.

4.2 Tasks to be Performed:

4.2.1 The Development of Job-Task Performance Tests for Job Activities Associated with Electronic Equipment. The contractor shall develop measurement procedures and appropriate scoring schemes for each of the following activities associated with the maintenance of electronic equipment: (1) performing equipment checkout procedures; (2) adjusting and aligning; (3) isolating between-stage faults to particular state (or functional unit or physically replaceable unit); (4) isolating within-stage faults to defective components, such as tubes, solid state devices, coil, capacitors or resistors; and (5) replacing of defective components. The contractor shall also be required to state his rationale for selection of the particular measurement procedure and scoring scheme for each type of activities. These procedures will be demonstrated by developing performance test appropriate scoring scheme for measuring the adequacy of each of the following AN/APN 147-AN/ASN 35 maintenance activities:

4.2.1.1 The entire checkout procedure for the AN/APN 147-AN/ASN 35 system (1 test).

4.2.1.2 All of the alignment procedures for this system (1 test).

4.2.1.3 Selected typical adjustments found in the system (5 tests). All adjustments should be identified and five (5) should be selected. These should be adjustments other than the routine ''front panel'' adjustments made during the checkout procedure.

4.2.1.4 Localization of between-stage faults to a particular stage (or functional unit or physically replaceable unit) (1 test with variable troubles). This testing procedure should be such that any defective unit could be inserted in the system and be identified as defective.

4.2.1.5 Isolation of within-stage fault to defective component (10 tests - one for each of 10 selected stages). All stages or functional units should be identified and the 10 most typical stages or functional units should be selected. Each test should be constructed so that different defective components can be inserted in the stage.

4.2.1.6 Removal and replacement of defective components (10 tests). Identify all types of component removal and replacement actions for the AN/APN 147-AN/ASN 35 system. Select ten (10) typical components removal and replacement operations for this system and develop a test for each of these operations.

4.2.2 The Development of Job-Task Performance Tests for Information Gathering Activities Associated with Electronic Equipment Maintenance. The contractor shall develop measurement procedures and appropriate scoring schemes for each of the following information gathering activities: (1) using oscilloscope to measure voltage and frequency, to compare waveshape standard and to make high accuracy time base measurements and comparisons; (2) using ohmmeter to measure direct-current resistance in electronic equipment; (3) using signal generator to inject standard or known signals to equipment for test purposes; (4) using tube checker to estimate quality of transistors. The test equipment normally used with the AN/APN 147-AN/ASN 35 system will be used as vehicles for developing the performance tests. However, the performance test developed for each test equipment will sample the full range of capabilities of each test equipment and will not be limited to only those capabilities required for maintenance of the

AN/APN 147-AN/ASN 35 system. Performance tests will be prepared for the following test equipment:

4.2.2.1 Tektronix Model 545A or B Oscilloscope

4.2.2.2 Model 505A - vacuum tube voltohmmeter

One test for testing of measuring voltage (all scales)

One test for measuring resistance (all scales)

4.2.2.3 Model WV 97 - RCA Electronic Voltohmmeter

One test for measuring voltage (all scales)

One test for measuring resistance (all scales)

4.2.2.4 Model SG-85/URM-25 RF Signal Generator

4.2.2.5 Model TS 382D/U Audio Oscillator

4.2.2.6 Model CMA 544A Test Set

4.2.2.7 Model CMA 546A Test Set

4.2.2.8 Model TS-148-Spectrum Analyzer

4.2.2.9 Model TV-2 Tube Checker

4.2.2.10 Model 1890M-Transistor Checker

4.2.3 <u>The Development of Job-Task Performance Tests for the Handtool Activities of Electronic Technicians</u>. The ultimate objective of this Air Force test development program will be to have testing procedures available for measuring most of the handtool operations required for use with most electronic equipments. This effort will, however, be limited to the handtool operations required to support the maintenance activities of the AN/APN 147-AN/ASN 35 system. The majority of the handtool operations are required to support the remove and replace activities. The contractor will be required to accomplish the following:

4.2.3.1 Identify each type of soldering task in the AN/APN 147-AN/ASN 35 system

4.2.3.2 Prepare a soldering exercise for each type of soldering task and to develop a scoring scheme reflecting the relative quality of performance of each exercise. Consideration should be given to the use of the R-F Probe in the measuring scheme. (See reference para. 3.1.1.7 above.)

4.2.3.3 Using the list of remove and replace tasks developed for 2.2.1.6 above, identify the handtools required for the mechanical portions of these activities. A test and scoring scheme should be developed to reflect the relative quality of performance for each of these handtools.

5.0 REPORTS AND DATA TO BE DELIVERED:

B001 - Progress Reports in accordance with Exhibit B, 19 Jun 1968.
B002 - Job-Task Performance Measurement Procedures for Electronic Maintenance Activities in accordance with Exhibit B, 19 Jun 68.
B003 - Technical Report in accordance with Exhibit B, 19 Jun 68.

6.0 SPECIAL CONSIDERATIONS:

6.1 Requirements for Contract Personnel: The principal investigator and the senior contract scientist must be mature individuals who had had successful experience in the development of job performance tests, which were used in controlled experiment(s) and reported in published document(s). Samples and descriptions of these tests must also be available for inspection. It will be necessary for these individuals to work on military sites in cooperation with operational personnel. It is, therefore, desirable that each of these individuals have had previous experience as electronic technicians in one of the military services. All contract personnel working on this research effort will require SECRET clearances.

6.2 Air Force Support

6.2.1 Government Furnished Property: The contractor will be furnished the following property as indicated within 30 days of the effective date of the contract.

6.2.1.1 The AN/APN-147 - AN/ASN-35 system.

This system will be available for the contractor's use at Wright-Patterson AFB, Ohio.

6.2.1.2 Applicable electronic test equipment.

This equipment will be available for the contractor's use at Wright-Patterson AFB, Ohio.

6.2.1.3 Copies of applicable maintenance technical orders.

These will be furnished for use at contractor facilities.

6.2.2 Air Force Personnel: The Air Force will furnish a maximum limit of three technicians as subjects for tryout of the performance tests. These subjects will be made available at an Air Force installation. The Air Force contract technical monitor will make appropriate arrangements for this tryout at the appropriate time. The arrangements also will include sufficient appropriate space for the tryout.

DO NOT FILM THIS PAGE