DOCUMENT RESUME

ED 087 471 IR 000 217

AUTHOR	Siegel, Arthur I.; And Others
TITLE	Adaptation of Advanced Measurement and Evaluation Techniques For Utilization in Air Force Technical Training Systems. Final Report.
INSTITUTION	Air Force Human Resources Lab., Lowry AFB, Colo. Technical Training Div.; Applied Psychological Services, Inc., Wayne, Pa.
REPORT NO	AFHRL-TR-73-18
PUB DATE	Nov 73
NOTE	152p.
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	Cluster Analysis; Educational Research; *Evaluation Techniques; Instructional Design; *Instructional Systems; Job Analysis; Job Skills; *Measurement Techniques; Military Training; Multidimensional Scaling; Program Evaluation; Systems Development; Task Analysis; *Technical Education; Testing; *Training
IDENTIFIERS	*Instructional Systems Development; ISD; U S Air Force

ABSTRACT

Research explored methods of providing a coherent framework for the Instructional Systems Development (ISD) model which would promote better system evaluation, internal communications and student testing. Lists of critical tasks for two Air Force specialties were compiled and multidimensional scaling and hierarchical cluster analysis employed to derive their respective job dimensions. A series of advanced measurement techniques was constructed on the basis of the job dimensions and then administered to groups of students. Results showed that: 1) multidimensional scaling analysis provided a useful framework for ISD models; 2) several of the advanced testing techniques were more useful predictors of student success than were the available multiple choice tests; and 3) student and instructor attitudes were positive. It was concluded that the scaling method should be used to order job analytic data and to provide coherency in ISD applications. In addition, the following new testing procedures were recommended as alternatives or adjuncts to the multiple choice format: sequential testing, figural systems, confidence testing, technical words, absurdity recognition, partial knowledge, and signal detection. (Author/PB)

AFHRL-TR-73-18

AIR FORCE

HUMAN RESOURCES

ED 087471

ADAPTATION OF ADVANCED MEASUREMENT AND EVALUATION TECHNIQUES FOR UTILIZATION IN AIR FORCE TECHNICAL TRAINING SYSTEMS

By

Arthur I. Siegel
Brian A. Berman
Applied Psychological Services, Inc.
Science Center
Wayne, Pennsylvania 19087

Gary G. Miller, Capt, USAF

TECHNICAL TRAINING DIVISION
Lowry Air Force Base, Colorado 80230

November 1973

Approved for public release; distribution unlimited.

LABORATORY

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

NOTICE

This final report was submitted by Applied Psychological Services, Science Center, Wayne, Pennsylvania, 19087, under contract F41609-72-C-0014, Task Number 03, with the Technical Training Division, Air Force Human Resources Laboratory, Lowry Air Force Base, Colorado, 80230. Captain Gary G. Miller was the contract monitor.

This report has been reviewed and is approved.

Marty R. Rockway, Technical Director
Technical Training 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-73-18	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ADAPTATION OF ADVANCED MEASUREMENT AND EVALUATION TECHNIQUES FOR UTILIZATION IN AIR FORCE TECHNICAL TRAINING SYSTEMS		5. TYPE OF REPORT & PERIOD COVERED Final 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Arthur I. Siegel Brian A. Berman Gary G. Miller		8. CONTRACT OR GRANT NUMBER(s) F41609-72-C-0014
9. PERFORMING ORGANIZATION NAME AND ADDRESS Applied Psychological Services, Inc. Science Center Wayne, Pennsylvania 19087		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 11210305
11. CONTROLLING OFFICE NAME AND ADDRESS Hq Air Force Human Resources Laboratory Brooks Air Force Base, Texas 78235		12. REPORT DATE November 1973 13. NUMBER OF PAGES 142
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Technical Training Division Air Force Human Resources Laboratory Lowry Air Force Base, Colorado 80230		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		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) technical training multidimensional scaling cluster analytic techniques measurement		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The multidimensional scaling and cluster analytic techniques were investigated as methods for providing a needed integrating framework within the course development and training evaluation context. Additionally, the relative merit was investigated of various advanced (novel) testing methods in the technical training context. The multidimensional and the cluster analytic techniques were held to provide the needed integrating thread and the advanced testing methods were indicated to possess advantage over the usual multiple choice examination.		

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

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SUMMARY

Problem

This study explored methods for providing a common definitional/orientational framework for the Instructional Systems Development (ISD) model so that each step in the ISD procedure could be based on a common set of integrating and coordinating concepts developed from the first step of the ISD procedure. Methods were also examined for integrating the various steps of the ISD procedure to allow for more systematic evaluation of ISD methods and improvement of communications among training program developers. The study also investigated the potential of certain advanced and novel testing techniques compared with the usual multiple choice tests currently being used in Air Force technical training. It was felt that certain novel testing techniques could provide for richer and more varied evaluative feedback that would aid in course and student evaluation.

Approach

In order to derive the basic job dimensions of the electronic principles and administrative specialist Air Force specialties, two lists of critical or frequently performed tasks were compiled from Occupational Survey Reports (ORSs) and Plans of Instruction (POIs). Multidimensional scaling and hierarchical cluster analysis were then employed to derive the job dimensions of the two specialties. The following advanced measurement techniques were constructed on the basis of the job dimensions derived from the preceding analyses: sequential, confidence, pictorial absurdities, pictorial, analogies, cognition of figural systems, partial knowledge, technical words, and Thurstone scaling. These advanced measurement techniques were administered to samples of electronics principles and administrative specialist students. Concurrent validity, reliability, cost/benefit, and attitudinal data were collected to evaluate the advanced measurement techniques.

Results

The results indicated that multidimensional scaling analysis of job tasks possesses considerable potential in ISD by providing a common framework for application of each step in the ISD procedure. Several of the advanced testing techniques seemed to be potentially more useful than the multiple choice tests being used, based upon prediction of end of block test scores. The attitudes of the Air Force technical training instructors and students enrolled in the two courses were, for the most part, favorable toward the advanced measurement techniques.

Conclusions

1. The multidimensional scaling method seemed to possess merit as a method for ordering job analytic data and for providing coherency within ISD application.

2. The advanced and novel testing procedures seemed to provide an adequate alternative or adjunct to the multiple choice format currently being used.

3. Psychometric, cost/effectiveness, and related properties favored consideration of the following advanced measures: sequential testing, figural systems, confidence testing, technical words, absurdity recognition, partial knowledge, and scoring on the basis of the theory of signal detection (d').

TABLE OF CONTENTS

Page

LIST OF TABLES

Table		Page

Table		Page

TABLE OF FIGURES

I. INTRODUCTION

Technical training within the Air Force is developed through a closed loop scheme known as the Instructional System Development (ISD) technique. The system is based on a systematic set of steps which aim to produce technical training programs which are maximally consonant with on-the-job requirements.

Within the ISD method, one step is concerned with test and evaluation. This test and evaluation aspect aims to develop measurement instruments/methods for determining whether or not the student has attained formalized criterion objectives.

The present program possessed two separate but interrelated purposes. The first purpose was to demonstrate methods for integrating the various steps in the ISD progression. The second purpose was to explore and demonstrate the utility of novel test concepts for student achievement measurement.

The ISD Model

Specifically, the revised ISD (AF Manual 50-2, 1970) model is composed of five formal steps:

1. Analysis of System Requirements--identification of the tasks to be performed within the overall environment of the operational system.

2. Definition of Educational or Training Requirements--determining the tasks that require instruction, the level of student proficiency to be developed, and the resources needed to conduct the instruction.

3. Development of Objectives and Tests--identification of the behaviors required for successful job performance and constructing criterion objectives and teaching steps, as well as achievement tests.

4. Planning, Development, and Validation of Instruction--selecting instructional methods, media, and equipment that best satisfy learning objectives, determining the sequencing of the instructional material, validation of instructional materials to insure that all elements of the instructional system function effectively in achieving stated objectives.

5. Conducting and Evaluating Instruction--identification of problem areas and corrective actions needed in order to satisfy the requirements of the operating commands.

It is possible, however, that if different persons with different points of view perform the various ISD steps, a certain amount of looseness might enter the instructional system. For example, the person developing the training requirements (step 2) might perceive the job analysis (step 1) differently from the original job analyst. And, the person developing the objectives (step 3) might perceive the requirements differently from the person who developed them, and so on. The result could be the introduction of considerable noise at various system nodes. This noise could reverberate and resonate throughout the entire system, with the end result that coherency is lost. What is needed is a method for ensuring system integrity. Specifically, the present report takes the point of view that if a set of integrating and coordinating constructs can be developed on the basis of the results of step 1, then these constructs can serve to define and focus each of the succeeding steps. The end result would be a common definitional/orientational framework throughout. Moreover, such a common framework would allow more systematic evaluation of the ISD methods and, more importantly, furnish a basis for communication among those involved in the training program development. For example, if the fifth step (conducting and evaluating instruction) is conducted in terms of criteria which are different from the training requirements (step 2), there can be no basis for discussion between the two groups performing the separate steps. The present report takes the point of view that multidimensional scaling analysis or cluster analysis of the task identified in step 1 (analysis of system requirements) can provide the needed common integrating core.

Multidimensional Scaling Analysis (MSA)

The two central problems in MSA are the determination of: (1) the minimum dimensionality of a given set of stimuli, and (2) the scale value of each stimulus on each of the dimensions. The specific experimental and computational procedures used have been described in detail by Torgerson (1952, 1958), Messick (1956a), and Messick and Abelson (1956).

The basic judgment on which the whole structure of MSA rests is very simple. In order to obtain estimates of the "psychological distances" among the various stimuli in the set, most experimenters have asked the subjects (judges) to indicate, in some manner, the degree of overall similarity between each stimulus pair. The methods for obtaining and scaling these similarity judgments are generally analogous to the classical psychophysical scaling techniques.

If the obtained scaled values can be taken as measures of the interstimulus distances in a Euclidean space, the analytical problem then becomes the determination of the number of axes in that space and the projections of the stimuli on these axes. In these final stages, MSA uses factor analysis. As in factor analysis, for example, the pattern of scale values (loadings) of the stimuli (tests) on each dimension (factor) presumably enables the experimenter to attach meaning to, and so to name, the dimensions.

There are a number of technical problems involved in MSA, such as the choice of method for obtaining the interstimulus distance estimates, the choice of spatial model to represent the distances, the determination of the constant required to set the distance estimates on a ratio scale (Messick & Abelson, 1956), and the decision as to whether a transformation of the basic data is required (Helm, Messick, & Tucker, 1961). Basically, however, MSA involves the steps of: (1) obtaining a matrix of interstimulus distances, and (2) determining the dimensionality of the space containing the stimulus points.

MSA also differs from certain other statistical techniques like factor analysis in that the results of the MSA grow out of the perceptions of the subjects who make the similarity judgments. The organization of the field that it produces is, therefore, the structure as perceived by these judges. What they consider correlated will be included in the emergent dimensions. Since the perceptions of the various types of judges may differ, the resulting dimensionality of the data may vary across different groups of judges.

MSA has been demonstrated to constitute a valid method for discovery of the underlying job dimensions in the job analytic context. For instance, Siegel and Smith (1965) used MSA to order the dimensions of the job of Civil Defense Director. Also, Schultz and Siegel (1962) and Siegel and Schultz (1963) performed a series of multidimensional scaling analyses in order to determine the job dimensions of Naval Aviation Electronics Technicians.

Hierarchical Clustering Scheme (HCS)

An alternative to MSA is some form of cluster analysis. Johnson (1967) demonstrated an algorithmic technique for grouping variables on the basis of empirical measures of similarity-dissimilarity. The technique, much like factor analysis, clusters homogeneous variables and is referrred to as a hierarchical clustering scheme (HCS). HCS has the inherent advantages of being rapid to compute manually, programmable for computer calculation, and open to several types of extensions or modifications.

The hierarchical features of the model are such that each subsequent cluster in an array of clusters is obtained by combining clusters at previous levels. Levels are determined by quantitative values which can be taken to indicate degrees of "strength." Eventually, all variables are clustered together to form the strongest clustering; the strongest cluster represents the merging of all previous clusters.

The clustering method produces solutions that are invariant under monotonic transformations of the data. Johnson states that the monotone transformation processes are dependent on the rank order of the data and that cluster analysis can be carried out knowing nothing of the data but the rank order. The values assigned to the clusterings are determined by rank order so that a monotonic transformation of the similarity matrix transforms the values of the clusterings, but leaves the clusterings invariant.

Advanced Measurement Techniques

As stated above, step 3 of the ISD technique involves the construction of measurement tests. Within the various Air Force technical training schools, these measurement tests often take the multiple choice format. While multiple choice achievement tests possess a number of advantages, there is reason to believe that other paper and pencil testing methods would provide a more varied and richer student evaluative basis. Accordingly, the second purpose of the present work involved exploration of some alternate test methods in the Air Force technical training context.

Additionally, the content areas covered by these alternate test approaches (called in this report "advanced" or "novel" measurement techniques) were drawn from, and based on, the results of the multidimensional scaling approach. Accordingly, the procedures serve as a test of the utility of the multidimensional scaling procedure for providing an integrating thread between the job analysis and the measurement testing.

Specific Purposes and Overview of Present Program

The specific purposes of the present program were to: (1) explore the utility of the multidimensional scaling/cluster analytic approaches as a basis for organizing certain aspects of the ISD procedures, and (2) investigate the utility of alternate test concepts in the technical training context.

To achieve these goals, two Air Force courses were selected for study--electronics principles and administrative specialist. These two courses represent wide differences in job content, training approach, and general required student aptitude. The MSA and the cluster analytic techniques were applied to available job analytic data for the technical specialties which graduates of these courses enter. Then, advanced measurement instruments were constructed on the basis of the extracted dimensions/clusters and the reliability, validity, and cost/effectiveness of these novel instruments were determined. The details of the methods involved and the results obtained are found in subsequent sections of this report.

II. METHODS

Development and Administration of Task List

It is important in multidimensional scaling, as in cluster analysis, to consider carefully the variables on which the analysis will be carried out, since the end result will reflect only the input data. For the present research, interest was centered on the behaviors involved in both the administrative specialist and electronics technician Air Force specialties. Therefore, two lists of behaviorally oriented job tasks were desired which would be inclusive of all the kinds of work performed by the men in these ratings, but which would not be so detailed as to require an impossibly large number of similarity comparisons or as to make the judgmental process unreasonably cumbersome.

With regard to the administrative specialist, the 35 tasks which were most critical or frequently performed by lower level airman administrative specialists were selected from relevant Occupational Survey Reports (OSR's). The tasks so selected were:

1. Type correspondence, directives, or reports
2. Prepare masters for reproduction
3. Extract information from files
4. Process and file correspondence
5. Post and insert changes in manuals, TO's, regulations, and similar publications
6. Operate office copying machines such as xerox, thermofax, or mimeograph
7. Prepare drafts of correspondence, directives or reports
8. Maintain active publication files
9. Maintain index to forms and publications
10. Maintain personnel locator files
11. Prepare and dispatch messages
12. Establish and maintain suspense files
13. Coordinate work activities with other sections or agencies
14. Schedule appointments and conferences
15. Maintain special order files
16. Prepare administrative orders

17. Requisition publications
18. Process outgoing mail
19. Maintain locator records
20. Distribute forms and publications
21. Receive, time stamp and route messages
22. Maintain inventory of forms and publications
23. Process official incoming mail for routing
24. Requisition supplies and equipment
25. Maintain status boards
26. Maintain duty rosters
27. Receive and process requisitions for forms and publications
28. Develop procedures for record maintenance and disposition
29. Prepare briefs of correspondence or reports
30. Maintain current routing guide and distribution lists
31. Develop and improve work methods and procedures
32. Operate key punch machine
33. Sort and distribute personal mail
34. Edit and review correspondence and reports
35. Determine requirements for equipment and supplies

A somewhat different approach was followed for the electronics specialist. The Occupational Survey Reports for electronics specialties were considered unsuitable because job experts at Keesler AFB indicated that these task listings did not adequately represent the Electronics Principles course. It was indicated that the Electronics Principles Plan of Instruction (POI), rather than the Electronics Specialist OSR's, should be used to develop the task list. Accordingly, Applied Psychological Services constructed a 26-item electronics principles task list from the Electronics Principles POI. The following items comprised the electronics principles list:

1. Knowing the purposes, uses, and application of components
2. Using equipment and tools properly and carefully
3. Cleaning up and maintaining orderly work area
4. Comparing the effects on circuit function of similar components
5. Describing components
6. Relating measures taken from components
7. Identifying electronic components and subclasses of components
8. Performing calculations
9. Identifying electronic components from schematics
10. Observing changes in measurements taken from electronic components
11. Inferring changes in component characteristics as the result of measurements
12. Measuring amplitude, capacitance, resistance, voltage, etc., with appropriate test equipments
13. Knowing the principles of circuit function
14. Knowing the characteristics of components
15. Knowing component limitations
16. Knowing the use, application, and limits of various circuit types
17. Knowing circuit theory
18. Identifying atomic components and their action
19. Knowing the theory of operation of components
20. Explaining effects on circuit function of electronic components
21. Using oscilloscope
22. Using miscellaneous test equipment
23. Troubleshooting from schematics
24. Reading schematics and identifying parts/components on schematics
25. Employing safety precautions for self
26. Employing safety precautions for equipment

Similarity Estimates

As a starting point for analysis, the multidimensional scaling model requires an estimate of the psychological distance between each pair of stimuli. Messick (1956b) has urged the multidimensional method of successive intervals as a simple, efficient, and desirable method for obtaining these data. He presents evidence and argues (1956c) that it: (1) takes less time than other methods, such as complete triads, (2) can therefore handle more stimuli in a given amount of time, and (3) produces equivalent results.

In the method of successive intervals or, as a special case, the method of equal appearing intervals, the distance judgments with regard to all possible stimulus pairs are indicated by the judge along a scale which is provided for him. For the present work, the stimulus material (either electronics or administrative, as appropriate) was presented in booklet form. At the top of each page in the booklet, one of the tasks was shown. Below it at the left side of the page, the remaining tasks were listed in a random order which was varied from one page to another. A scale running from 1 to 11 appeared to the right of each of the tasks. The scale points1 and 2 were described at the top of the page as representing a judgment of "very similar"; points 3, 4, and 5, as representing "moderately similar"; points 7, 8, and 9, as representing "moderately different"; and points 10 and 11, as representing "very different." Scale point 6, in the middle of the range, was not described in verbal terms. A sample page of the administrative specialist form, called the Technical Task Inventory, is shown in Table 1.

The directions asked the subject to compare each task listed with the one shown at the top of the page and then to "indicate by check in the appropriate column to the right how similar or different the two tasks are." The complete cover page of the form, including the directions, is shown in Table 2.

Sample Similarity Estimate Page

Maintain Index To Forms And Publications	Very Similar 1	2	Moderately Similar 3	4	5	6	Moderately Different 7	8	9	Very Different 10	11
Post and insert changes in manuals, TO's, regulations, and similar publications											
Prepare and dispatch messages											
Schedule appointments and conferences											
Develop procedures for record maintenance and disposition											
Maintain special order files											
Type correspondence, directives, or reports											
Requisition publications											
Requisition supplies and equipment											
Establish and maintain suspense files											
Develop and improve work methods and procedures											
Maintain duty rosters											
Process outgoing mail											
Prepare briefs of correspondence or reports											
Maintain current routing guide and distribution lists											
Extract information from files											
Prepare drafts of correspondence, directives, or reports											
Operate key punch machine											
Maintain personnel locator files											
Maintain locator records											
Distribute forms and publications											
Maintain status boards											
Process official incoming mail for routing											
Prepare administrative orders											
Receive time stamp and route messages											
Process and file correspondence											
Prepare masters for reproduction											
Receive and process requisitions for forms and publications											
Edit and review correspondence and reports											
Maintain inventory of forms and publications											
Sort and distribute personal mail											

Table 2

Directions for Similarity Estimates

Name ______________________ Today's Date ______________

TECHNICAL TASK INVENTORY

The purpose of this Inventory is to compare various tasks performed by Administrative Specialists. Each task is shown in a box at the top of each page. Below it is a list of other tasks. You should compare each task in the list with the one in the box at the top of the list and indicate by a check in the appropriate column to the right how similar or different the two tasks are. There are no "right" or "wrong" answers to this inventory; your best judgments of similarity are the only "right" answers.

Before you begin, open the booklet and look over the pages briefly to get an idea of what tasks are included. Notice that the pages have different numbers of tasks listed. Some pages have two or more separate lists, each with its own boxed comparison item.

Start working at the beginning of the booklet. Try to vary your check marks so that some appear in all eleven columns. Do not hesitate to use the extreme responses numbered 1 and 11, if you feel any comparison deserves one of them.

EXAMPLE

Maintain Training Records

	Very Similar 1	2	Moderately Similar 3	4	5	6	Moderately Different 7	8	9	Very Different 10	11
Prepare records for controlled mail........	__	__	✓	__	__	__	__	__	__	__	__
Assign specific work to individuals........	__	__	__	__	__	__	__	__	__	✓	__

The first check indicates that the person completing the inventory thinks that "prepare records for controlled mail" is moderately similar to "maintain training records."

WHEN YOU HAVE FINISHED, CHECK BACK TO MAKE CERTAIN YOU HAVE PLACED A CHECK NEXT TO EACH TASK IN THE LIST ON EVERY PAGE.

In addition to the random order of tasks on each page, the order of pages in each booklet was randomized. Each subject, then, completed a form with differently ordered pages. This control attempted to avoid contamination from order, set, and fatigue effects.

Twenty instructors in the administrative specialist course at Keesler AFB completed the administrative specialist questionnaire, while 29 electronic instructors completed the electronics principles questionnaire. Separate sessions were involved for each group.

A very brief, general description of the program was given at the beginning of each session. The booklets were essentially self-administering. Almost all of the instructors completed the forms within two hours.

The subjects were able to understand their task easily. Most of them proceeded without difficulty and with almost no questions.

The Matrices of Interstimulus Distances

For each specialty (administrative or electronics) the scale value for each pair of job tasks was taken as the mean of the values checked along the similarity scale by the judges in the appropriate group. The obtained scale values, or intertask distances, for the electronics principles instructors and for the administrative specialist instructors are presented in Tables 3 and 4, respectively.

Table 3

Mean Original Intertask Distances for Electronics Principles

Task	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	6.24	8.86	3.41	2.83	3.68	3.93	5.66	4.17	4.38	5.00	4.83	2.83	2.48	2.48	2.07	2.07	5.97	2.48	3.34	5.34	6.66	4.10	3.55	7.55	5.41
2		3.07	7.79	7.79	6.55	9.07	7.55	8.72	4.03	6.72	3.00	6.34	9.03	6.62	5.59	7.62	9.52	8.93	7.86	3.03	2.41	6.55	8.90	2.24	2.03
3			9.48	9.86	8.45	9.62	9.90	9.69	8.03	9.45	7.17	8.36	9.72	8.97	9.00	9.28	10.00	8.97	9.76	7.24	7.14	9.48	9.86	3.52	4.21
4				4.41	3.07	5.24	5.24	4.69	3.31	4.62	5.00	3.14	2.69	4.41	3.31	3.14	7.45	2.79	3.28	4.34	6.00	5.38	4.86	8.43	8.14
5					6.31	3.76	6.59	3.97	5.00	5.41	5.07	3.69	2.34	3.66	3.83	3.45	6.00	2.72	3.03	7.17	7.11	6.97	3.66	9.31	7.17
6						6.17	3.97	6.28	2.21	3.31	3.28	4.41	3.03	4.03	4.66	5.03	7.76	4.89	4.62	3.48	5.62	5.52	7.31	8.36	7.72
7							7.62	2.21	5.97	5.31	4.28	6.17	5.90	5.24	3.96	2.72	5.79	5.38	4.03	7.17	5.83	2.03	1.55	10.38	8.00
8								7.64	4.59	4.21	3.45	5.62	5.63	5.28	5.03	4.38	6.86	5.83	4.79	4.28	4.72	4.45	8.45	9.62	8.28
9									6.24	5.69	6.90	5.83	4.14	6.00	5.76	6.10	6.24	4.66	4.83	6.69	8.28	5.90	2.45	9.52	9.31
10										3.28	2.52	4.69	3.86	4.21	4.10	4.48	7.62	4.07	4.10	2.62	5.17	5.17	7.90	6.52	6.48
11											3.62	3.17	3.10	3.45	4.07	4.17	7.31	2.76	3.07	4.79	4.07	5.14	6.90	8.03	7.03
12												3.90	3.59	4.79	4.10	5.45	7.66	4.72	4.82	1.79	1.45	5.07	7.21	6.00	5.69
13													3.76	3.62	1.72	1.55	5.24	2.93	2.14	4.41	4.52	2.62	5.00	7.45	6.45
14														2.38	2.59	2.31	4.07	2.66	3.69	4.93	4.38	3.14	5.97	9.14	6.34
15															2.48	2.97	6.45	2.97	4.10	5.28	6.41	4.54	5.38	7.38	7.10
16																2.00	8.72	2.28	2.90	5.28	6.69	3.93	5.45	6.93	7.52
17																	6.24	1.69	2.21	5.21	6.62	3.59	5.75	8.11	8.21
18																		3.62	5.89	8.93	7.69	5.66	8.52	8.90	8.66
19																			2.66	5.41	4.90	2.72	4.86	7.00	6.76
20																				5.14	5.31	2.55	5.79	8.69	8.07
21																					3.14	6.07	6.83	5.62	4.62
22																						5.17	8.69	5.31	4.45
23																							2.83	8.07	8.03
24																								9.45	9.34
25																									2.75

Table 4

Mean Original Intertask Distances for Administrative Speci

Task	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	3.45	6.70	5.40	8.00	7.75	2.05	8.20	7.85	9.55	3.60	7.65	6.60	7.95	7.30	2.70	6.05	8.65	8.80	8.50	8.25	7.80	8.60	7.25
2		8.25	8.00	9.15	6.25	3.00	7.60	8.15	9.35	6.75	9.05	8.35	10.00	8.85	4.15	8.60	9.75	8.90	9.55	9.75	9.30	9.40	8.15
3			3.45	8.05	9.20	5.80	7.20	8.60	7.05	7.65	5.80	9.00	8.45	4.50	6.30	7.10	8.40	7.95	8.50	7.95	6.65	7.50	8.00
4				6.85	9.00	5.55	6.75	8.20	6.20	5.25	4.15	7.50	7.75	4.05	6.70	7.65	5.55	5.35	6.45	5.25	7.15	4.70	7.60
5					9.80	8.30	2.25	3.05	7.30	8.05	6.00	8.85	9.95	6.40	7.90	6.30	9.40	8.05	6.75	8.30	6.10	8.30	8.70
6						8.85	10.00	9.80	9.75	8.95	10.05	9.50	9.55	9.70	8.80	9.90	9.85	9.75	10.05	9.05	9.80	9.85	9.20
7							8.80	8.80	8.35	3.00	7.80	7.90	8.40	8.85	3.90	7.35	8.60	8.75	9.05	8.05	8.20	8.60	7.35
8								1.90	5.15	8.95	5.45	7.50	9.00	4.45	8.50	3.30	8.30	6.40	6.25	9.10	4.10	7.80	7.25
9									7.10	8.30	5.45	8.95	8.80	4.55	9.30	4.90	9.25	6.80	7.10	8.90	4.20	8.95	7.00
10										8.35	5.05	6.65	8.60	5.50	8.70	9.30	6.80	2.40	7.95	8.45	6.05	6.90	9.25
11											7.40	7.50	8.85	8.65	6.10	7.45	4.65	9.40	7.90	3.55	8.70	4.25	7.95
12												7.85	6.40	4.55	6.95	5.50	6.60	3.95	5.60	6.15	5.50	6.25	7.50
13													4.40	8.85	8.35	7.65	7.60	7.50	7.20	8.60	9.05	6.15	6.55
14														9.50	8.95	9.50	8.85	8.30	9.75	8.65	8.75	8.65	9.25
15															6.20	8.20	8.10	5.15	8.60	8.65	6.20	8.55	8.20
16																7.30	8.95	8.95	9.05	8.95	8.45	9.15	7.85
17																	7.50	8.35	4.65	7.45	4.40	7.50	3.00
18																		7.95	4.25	3.60	8.10	2.35	8.25
19																			8.10	8.05	6.50	6.60	8.55
20																				6.15	4.65	4.20	6.55
21																					7.35	3.20	8.25
22																						7.95	5.60
23																							8.65
24																							
25																							
26																							
27																							
28																							
29																							
30																							
31																							
32																							
33																							
34																							

Multidimensional Scaling Analysis

The methods of multidimensional scaling analysis have been fully described in a number of sources (e.g., Shepard, et al., 1972). These standard methods were followed in the present analysis. Often, in MSA, an additive constant is computed which when added to all the relative intertask distances converts them to absolute distances (Messick & Abelson, 1956). However, prior MSA studies with job task lists (Schultz & Siegel, 1962, 1963; Siegel & Schultz, 1963; Siegel & Smith, 1965) have indicated that, at least for the job analytic context, the constant is minimal and the correction exerts no effect on the dimensionality of the results.

The similarity estimates, as obtained, were first reversed in directionality so that increased similarity was indicated by higher numbers. The resulting scale values were converted to proportions of one and the resultant values were placed in similarity ("distance") matrix form. Each of the matrices was factored by the method of principle components. The resultant unrotated solutions were then rotated to orthogonal simple structure as determined by the varimax criterion (Kaiser, 1958). The factor loadings derived from a MSA must be interpreted somewhat differently than those which result from a factor analysis, because the data input consist of similarity estimates rather than correlation coefficients. Each factor loading in the present case can be considered to represent the similarity between a task and a particular factor.

In factor analysis, an item is considered, by convention, to be meaningfully associated with a factor if it correlates .3 or better with the factor. With regard to our data, though, a value of .3 would indicate moderate dissimilarity with the factor; therefore, a value of .5 was adopted as the acceptable criterion for item acceptance into a factor or dimension.

The results of the factor analyses resulted in the identification of nine orthogonal factors as representing the 35 administrative specialist tasks and four factors for the 26 electronics principles tasks.

Dimensions of Electronics Principles Tasks

The stimulus task functions with the highest loadings on the first of the four factors extracted from the electronics principles task list are presented in Table 5. The task functions in factor 1 are primarily concerned with those knowledges concerned with circuit and component theory and function. Hence, factor 1 was named "Component and Circuit Knowledge."

Table 5

Stimulus Task Functions Possessing the Highest Loadings on Factor 1, Component and Circuit Characteristics

Task Function Number	Task Function	Loading
17	Knowing circuit theory	.848
13	Knowing the principles of circuit function	.823
20	Explaining effects on circuit function of electronic components	.776
16	Knowing the use, application, and limits of various circuit types	.727
19	Knowing the theory and operation of components	.672
1	Knowing the purposes, uses, and application of components	.609
5	Describing components	.605
4	Comparing the effects on circuit function of similar components	.601
14	Knowing the characteristics of components	.567
15	Knowing component limitations	.535
23	Troubleshooting from schematics	.510

The electronic task functions loading heavily on factor 2 are presented in Table 6. Three of the four tasks are concerned directly with safety precautions. Accordingly, factor 2 was named "Safety."

Table 6

Stimulus Task Functions Possessing the Highest Loadings on Factor 2, Safety

Task Function Number	Task Function	Loading
25	Employing safety precautions for self	.928
26	Employing safety precautions for equipment	.862
2	Using equipment and tools properly and carefully	.840
3	Cleaning up and maintaining orderly work area	.725

The stimulus task functions with the highest loadings on factor 3 are described in Table 7. These task functions are mainly concerned with the use of test equipment; therefore, this factor was called "Testing."

Table 7

Stimulus Task Functions Possessing the Highest Loadings on Factor 3, Testing

Task Function Number	Task Function	Loading
22	Using miscellaneous test equipment	-.788
12	Measuring amplitude, capacitance, resistance, voltage, etc., with appropriate test equipments	-.730

Two weak factors with similar content were combined and are called factor 4. The stimulus task functions with the heaviest loadings on the combined factor 4 relate to reading schematics in order to identify electronic components. This final electronics factor was named "Component Identification." The items with the highest loadings on factor 4 are presented in Table 8.

Table 8

Stimulus Task Functions Possessing the Highest Loadings on Factor 4, Component Identification

Task Function Number	Task Function	Loading
24	Reading schematics and identifying parts/components on schematics	-.936
9	Identifying electronic components from schematics	-.712
.....	.....	.....
7	Identifying electronic components and subclasses of components	.924
9	Identifying electronic components from schematics	.586

Dimensions of Administrative Specialist Tasks

For the administrative specialist, the stimulus task functions, loading heavily on factor 1, are presented in Table 9. The task functions with the highest loadings on factor 1 seem to represent functions of the job that relate to the preparation of documents and other communications. Accordingly, this factor was named "Document Preparation."

Table 9

Stimulus Task Functions Possessing the Highest Loadings on Factor 1, Document Preparation

Task Function Number	Task Function	Loading
1	Type correspondence, directives or reports	.896
2	Prepare masters for reproduction	.760
7	Prepare drafts of correspondence, directives or reports	.894
16	Prepare administrative orders	.768
29	Prepare briefs of correspondence or reports	.785
11	Prepare and dispatch messages	.597
34	Edit and review correspondence and reports	.531

The stimulus task functions with the highest loadings on factor 2 represent job behaviors involved in the handling of mail and other communications. This factor was named "Communications Processing."

Table 10

Stimulus Task Functions Possessing the Highest Loadings on Factor 2, Communication Processing

Task Function Number	Task Function	Loading
23	Process official incoming mail for routing	.846
18	Process outgoing mail	.837
33	Sort and distribute personal mail	.755
21	Receive, time stamp, and route messages	.803
11	Prepare and dispatch messages	.697

The stimulus task functions loading most heavily on factor 3 are shown in Table 11. The tasks with the highest loadings on factor 3 relate to updating and indexing publications. Accordingly, factor 3 was named "Publication Maintenance." All significant loadings on this factor were negative.

Table 11

Stimulus Task Functions Possessing the Highest Loadings on Factor 3, Publication Maintenance

Task Function Number	Task Function	Loading
8	Maintain active publication files	-.882
9	Maintain index to forms and publications	-.856
5	Post and insert change in manuals, TO's, regulations, and similar publications	-.857

The stimulus task functions with the highest loadings on factor 4 are presented in Table 12. The tasks loading on factor 4 are concerned with record keeping and filing. Accordingly, factor 4 was named "File Maintenance."

Table 12

Stimulus Task Functions Possessing the Highest Loadings on Factor 4, Fire Maintenance

Task Function Number	Task Function	Loading
19	Maintain locator records	.892
10	Maintain personnel locator files	.793

The administrative specialist stimulus task functions with the highest loadings on factor 5 are presented in Table 13. The tasks represented by factor 5 relate to requisitioning and processing associated with the acquisition of supplies and equipment. Accordingly, this factor was named "Supply and Equipment Processing."

Table 13

Stimulus Task Functions Possessing the Highest Loadings on Factor 5, Supply and Equipment Processing

Task Function Number	Task Function	Loading
24	Requisition supplies and equipment	-.896
35	Determine requirements for equipment and supplies	-.851
27	Receive and process requisitions for forms and publications	-.570

The stimulus task functions loading most heavily on factor 6 are presented in Table 14. The tasks with the highest loadings on factor 6 use action words which relate to filing in the active sense, i.e., "process," "extract," and "maintain." This factor, like factor 3, contained no tasks with significant positive loadings. Factor 6 was named "Filing."

Table 14

Stimulus Task Functions Possessing the Highest Loadings on Factor 6, Filing

Task Function Number	Task Function	Loading
3	Extract information from files	.822
4	Process and file correspondence	.590
15	Maintain special order files	.616

The stimulus tasks with the heaviest loadings on factor 7 are presented in Table 15. The tasks loading on factor 7 refer to the development of clerical methods and procedures. Hence, factor 7 was named "Job Structuring and Development."

Table 15

Stimulus Task Functions Possessing the Highest Loadings on Factor 7, Job Structuring and Development

Task Function Number	Task Function	Loading
28	Develop procedures for record maintenance and disposition	.874
31	Develop and improve work methods and procedures	.763

The stimulus tasks with the highest loadings on factor 8 are presented in Table 16. This factor is mainly concerned with the maintenance of lists of various kinds. Accordingly, factor 8 was named "List Maintenance."

Table 16

Stimulus Task Functions Possessing the Highest Loadings on Factor 8, List Maintenance

Task Function Number	Task Function	Loading
25	Maintain status boards	.806
30	Maintain current routing guide and distribution lists	.513

Finally, factor 9 relates to the operation of office machinery and was called "Machine Operation." The tasks with the highest loadings on factor 9 are presented in Table 17.

Table 17

Stimulus Task Functions Possessing the Highest Loadings on Factor 9, Machine Operation

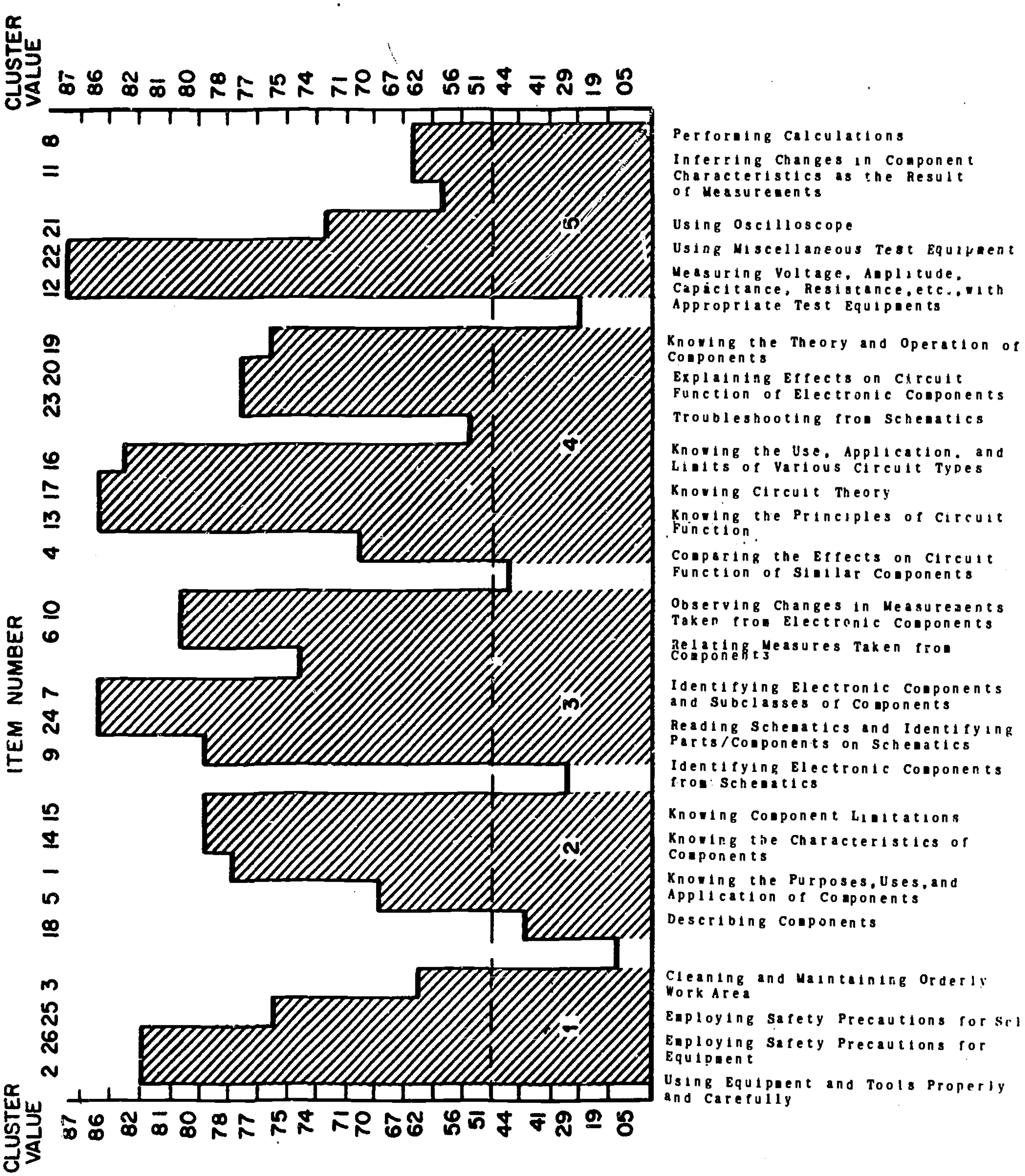
Task Function Number	Task Function	Loading
6	Operate office copying machines such as xerox, thermofax, or mimeograph	.879
32	Operate keypunch machine	.805

Cluster Analysis

As a check on the factor analysis and in order to explore the merit of cluster analysis for job dimensionality determination, a hierarchical cluster analysis was performed on the similarity estimates for the separate groups. Cluster analysis is a method which is ideally suited to the reduction of intertask distance matrices. No special assumptions are required for use of the method, and it can be applied to almost any kind of distance or correlational data. It has the added advantage of simplicity and ease of interpretation. A matrix reduction can easily be accomplished by hand if the variables number 25 or less. If large matrices are involved, a computer program is available which requires only seconds of computer time to arrive at a solution. The results of cluster analysis automatically produce "simple structure." The method is totally insensitive to data transformation, a fact which avoids the additive constant derivation.

The method of Johnson (1967), described previously, was used in the present analysis. Table 18 presents the cluster analytic results for the electronic principles tasks. The dotted line across Table 18 is the cutoff value for acceptance of an item into a cluster. Each stalk or bar in the histogram, except for these connecting clusters, represents an item. The item numbers are written across the top of the histogram and the clusters

ELECTRONIC PRINCIPLES CLUSTER ANALYSIS



are included at the bottom. The cluster values are shown at the sides of the histogram. Table 19 presents the name given to each cluster shown in Table 18.

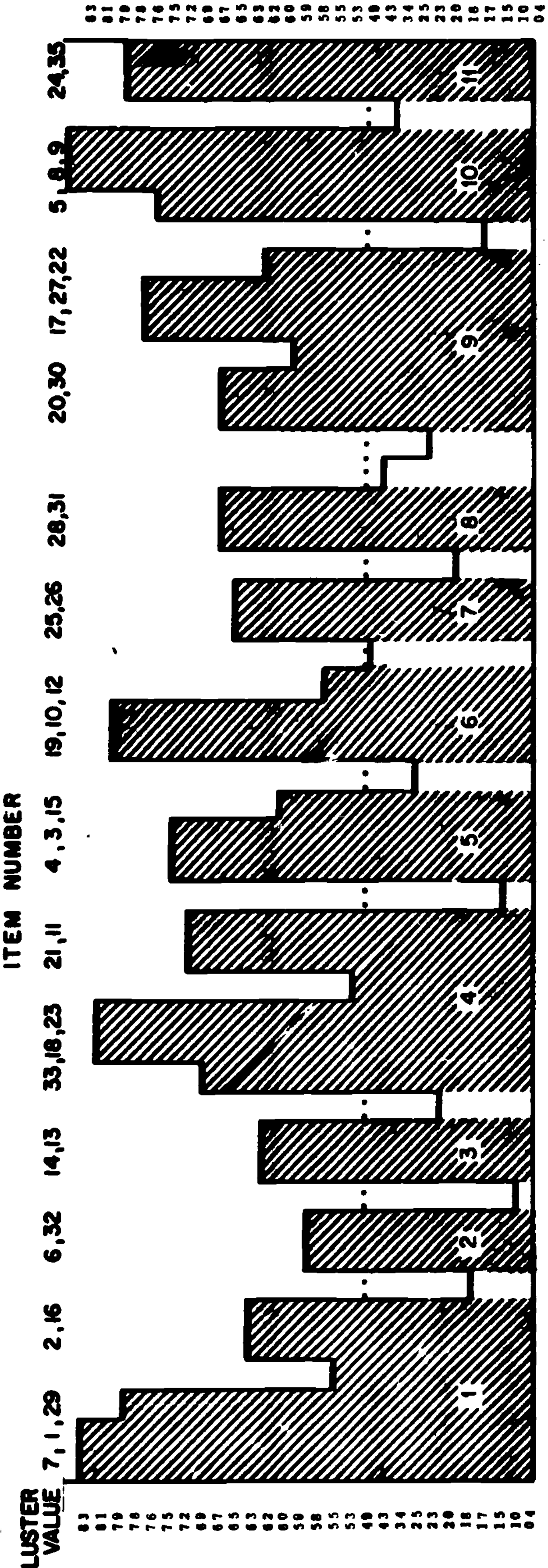
Table 19

Cluster Names for the Electronics Principles Clusters Analysis

Cluster Number	Cluster Name
1	Safety
2	Component Characteristics
3	Component Identification and Measurement
4	Component and Circuit Characteristics
5	Electronic Analysis

Similarly, Tables 20 and 21 present the cluster analytic results and cluster names for the administrative specialist questionnaire. For both cluster analyses, the item cutoff point was set as close to .5 as possible, because items that cluster below .5 are more dissimilar than similar to each other, while items that cluster above .5 are more similar than dissimilar to each other.

TABLE 20
ADMINISTRATIVE SPECIALIST CLUSTER ANALYSIS



Determine Requirements for Equipment and Supplies
Requisition Supplies and Equipment

Maintain Index to Forms and Publications
Maintain Active Publication Files
Post and Insert Changes in Manuals, TO's, Regulations, and Similar Publications

Maintain Inventory of Forms and Publications
Receive and Process Requisitions for Forms and Publicatio
Requisition Publications

Maintain Current Routing Guide and Distributions List
Distribute Forms and Publications

Develop and Improve Work Methods and Procedures
Develop Procedures for Record Maintenance and Disposition

Maintain Duty Rosters
Maintain Status Boards

Establish and Maintain Suspense Files
Maintain Personnel Locator Files
Maintain Locator Records

Maintain Special Order Files
Extract Information from Files
Process and File Correspondence

Prepare and Dispatch Messages
Receive Time Stamp and Route Messages

Process Official Incoming Mail for Routing
Process Outgoing Mail
Sort and Distribute Personal Mail

Coordinate Work Activities with Other Sections or Agencies
Schedule Appointments and Conferences

Operate Keypunch Machine
Operate Office Copying Machines such as Xerox, Thermofax, or Mimeograph

Prepare Administrative Orders
Prepare Masters for Reproduction

Prepare Briefs of Correspondence or Reports
Type Correspondence, Directives, or Reports
Prepare Drafts of Correspondence, Directives, or Reports

Table 21

Cluster Names for the Administrative Specialist Clusters Analysis

Cluster Number	Cluster Name
1	Document Preparation
2	Machine Operation
3	Job Structuring
4	Communication Processing
5	Filing
6	File Maintenance
7	Board Maintenance
8	Development of Clerical Methods and Procedures
9	Form and Publication Processing
10	Publication Maintenance
11	Supply and Equipment Maintenance

Algorithmic Integration--Electronic Principles

Since the multidimensional scaling and the cluster analytic algorithms are based on entirely different suppositions, any similarity of transmethod results is evidence for the robustness of the current solutions and for the underlying dimensionality of the task data.

The results of the cluster and the factor analytic algorithms, when applied to the electronic principles interstimulus distance matrix, were not as strikingly similar as the results derived from the administrative specialist interstimulus distance matrix, described below.

Comparison of the electronic principles results across methods revealed that clusters 2 and 4 are totally contained in factor 1. Apparently, the cluster analysis, in this case, differentiated to a greater degree and separated component characteristics from circuit characteristics.

The stimuli that loaded the highest on factor 3 (Safety) are exactly the same as the stimuli in cluster 1 (Safety). On the other hand, only two of the items in cluster 5 (Electronic Analysis) loaded on factor 3 (Testing).

Finally, the three items in factor 4 (Component Identification) are also members of cluster 3 (Component Identification and Measurement).

Algorithmic Integration--Administrative Specialist

The similarity of results across both techniques is quite striking for the administrative specialist task data.

The five items of cluster 1 (Document Preparation) are also contained in factor 1 (Document Preparation). We note that the two items (item 11 and item 34) that failed to cluster in cluster 1 had the lowest factor loadings on that factor (.597 and .531 respectively). The five items that are common to both algorithms, though, had loadings of .760 or higher. The items making up factors 2, 3, 6, 7, and 9 are exactly the same as the items in clusters 4, 10, 5, 8, and 2 respectively. This means that almost perfect content congruency was exhibited for six of the nine factors.

Two items in factor 4 (File Maintenance) were the same as two of the three items in cluster 6 (File Maintenance). The third item in cluster 6 (item 12) had a cluster value of .54, while the two common items clustered at .78. As with factor 1, it seems as though those items that failed to fall into the same cluster/factor had the lowest loading value with that cluster or factor.

Similar results were obtained for factor 5 and cluster 11. Two items in cluster 11 (Supply and Equipment Maintenance) were the same as two of the three items in factor 5 (Supply and Equipment Maintenance). The one item in factor 5 (item 27) which did not cluster with cluster 11 had a loading of -.570 on factor 5.

Finally, factor 8 seems to be the only factor which did not correspond to any cluster. The two items in factor 8 (items 25 and 30) fell into different clusters. Also, clusters 3, 7, and 9 had no corresponding factors.

Accordingly, it seems that the cluster analytic solution produced greater differentiation than the factor analytic approach. However, it seems that the cluster analytic approach may have overfragmented the tasks to produce a result which would be obtained from overfactoring. However, these results warrant the conclusion that both solutions yielded highly similar structure for the administrative specialist interstimulus distance matrix.

These results lead to the conclusion that, in the task taxonomic context, the differences between the results yielded by the two methods seem to be that cluster analysis may tend to produce more than one cluster for each analytically derived factor. Nevertheless, the results from both methods are similar enough to support the contention that a suitable descriptive structure can be obtained from either technique.

Basis for Test Selection

As stated earlier, for the tasks performed by the administrative specialist and the electronics technician in the Air Force, a multidimensional scaling analysis was performed. Both factor analytic and hierarchical cluster analytic procedures were applied to yield the basic organizational structure. The results of the application of both analytic techniques seemed largely congruent within both specialties. However, the factor analytic results were adopted for the advanced test methods developmental aspects of the present work.

A separate analysis was performed to yield a list of the types of advanced measurement techniques which seemed of value in the Air Force. This analysis was largely drawn on a previously developed report (Siegel, Bergman, & Sellman, 1972).

This test list was then employed as one axis of a matrix, while the factorial structure of the task list was employed as the second axis. For each test type, a factor assignment was made. The results for electronics are presented in Table 22. (The safety factor was eliminated from Table 22 inasmuch as students enrolled in Block IV of the Electronic Principles course, the course Block from which students were subsequently drawn for test tryout purposes, are minimally responsible for knowledge of the various safety procedures.) As indicated in Table 22, it was our goal to measure each factor (up to the completion of Block IV) by a separate test type. Accordingly, there was a confidence testing procedure applied to measurement of the component circuit characteristics course factor, a pictorial test of the component identification factor, etc. The need achievement test was not factorially related and the psychophysiological reaction was obtained in conjunction with the confidence test administration. Scoring in accordance with theory of signal detection took place in conjunction with the confidence test.

Table 22

Test Type for Testing Each Factor in Electronics Principles Course

Test Type	Course Factor		
	Component/Circuit Characteristics	Testing	Component Identification
Sequential Testing	✓		
Confidence Testing	✓		
Pictorial Testing			✓[3]
Cognition of Figural Systems		✓(symbolic)	
Analogies		✓	
Signal Detection[1]	✓		
Need Achievement[2]			
Psychophysiological Arousal[1, 2]			

[1] Administered in conjunction with confidence testing
[2] Not factorially related
[3] Pictorial schematic reading and pictorial absurdities

The same logic was applied to the administrative specialist, and the results are presented in Table 23. The various advanced techniques are all applied to the same work factor in this case because Block II of the course, the block selected as a basis for the current work, largely involves learning typewriting within the document preparation context.

Electronics Tests

The electronics advanced measurement techniques included: (a) sequential testing, (b) confidence testing, (c) pictorial absurdities, (d) pictorial schematic reading, (e) configuration of figural systems, (f) analogies, (g) signal detection, (h) a measure of need achievement, and (i) a measure of psychophysiological arousal.

The development and use of each of these testing techniques (except need achievement and psychophysiological arousal) is described in detail in the following paragraphs. The need achievement and psychophysiological arousal methods, which were also employed for the administrative testing, are discussed at the conclusion of the description of the various advanced measurement techniques.

Sequential Testing

Sequential testing involves a branching sequence in which the student receives items at a difficulty level that is appropriate for his level of accomplishment. If a student fails to reach criteria on his first block of test items, he is routed to an easier block of items. If, on the other hand, he exceeds criterion on the first block of items, he is routed to a more difficult block of items. This procedure has the following advantages: (1) it decreases test time (especially for students at the extremes of the distribution because they can be routed quickly), (2) it increases reliability, and (3) it increases student motivation because the student is not forced to take and guess at more difficult items.

Construction of sequential tests requires determination of item difficulty. Since the items within the tests currently employed in the electronic principles Block IV have been extensively item analyzed, these items were used as the basis of the sequential test.

Table 23

Test Type for Testing Each Factor in Administrative Specialist Job

	Job Factor								
	Document Preparation	Communications Processing	Publications Maintenance	File Maintenance	Supply/ Equipment Maintenance	Filing	Job Structuring	List Maintenance	Machine Operation
Partial Knowledge	✓								
Technical Words	✓								
Pictorial	✓ (absurdities)								
Thurstone Scaled Typing Samples	✓			NOT IN BLOCK II					
Signal Detection	✓								
Need Achievement	a								
Psychophysiological Arousal	a, b								

a = not factorially related.

b = administered in conjunction with partial knowledge test.

Five nonoverlapping tests of 12 multiple choice items each were selected from the 200 items used in Block IV of the electronic principles course. The sequential subtest number, difficulty range (the range of difficulty of the various Block IV items selected for inclusion in a particular sequential subtest), and mean difficulty value (the proportion of persons in an Air Force pool of electronics students responding correctly to the items) for the five, five minute tests developed from these items are presented in Table 24.

Table 24

Sequential Number, Difficulty Range, and Mean Difficulty Value for Each Sequential Subtest

Subtest	Difficulty Range	Mean Difficulty
1	.89 - .95	.92
2	.77 - .83	.80
3	.65 - .71	.68
4	.53 - .59	.56
5	.41 - .47	.45

Sample items from sequential test 3 are shown in Figure 1. In order to determine if the adjacent sequential tests were significantly different from each other, an analysis of variance was performed. This analysis compares test 1 results with test 2 results, test 2 results with test 3 results, etc. The total N in the analysis was 60, the number of items making up the sequential test battery, because the intent of the analysis was to compare difficulty values across subtests. The results of this analysis are summarized in Table 25.

1. FREQUENCY DISTORTION CAN BE CAUSED BY

   a. linear operation

   b. non-linear operation

   c. removing bypass capacitors

   d. reactive components

2. POWER DISSIPATION IN A TRANSISTOR CAUSES AN INCREASE IN

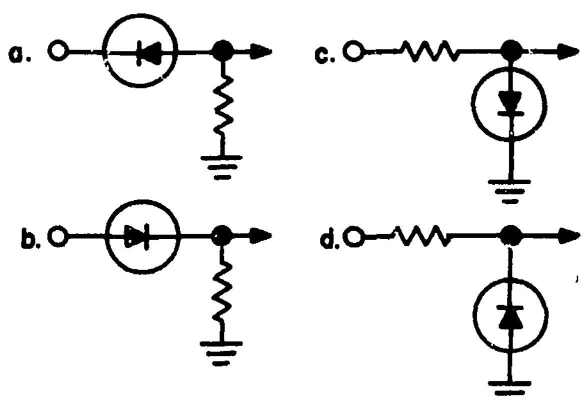
   a. voltage gain

   b. junction resistance

   c. junction temperature

   d. the forward current transfer ratio

3. SELECT THE SCHEMATIC DIAGRAM OF A SHUNT NEGATIVE LIMITER



4. THE PLACEMENT OF C1 MAKES THE FILTER A

   a. capacitor input L type

   b, capacitance type

   c. L type RC

   d. L type LC

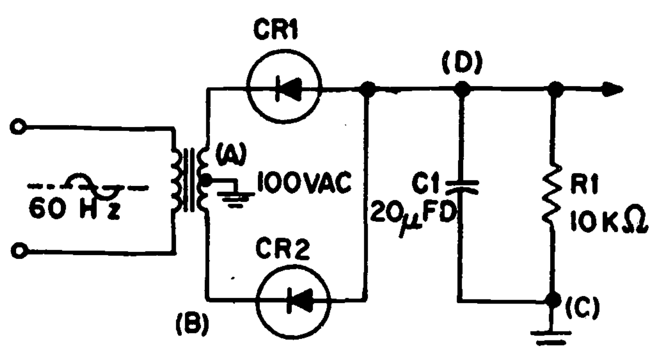


Figure 1. Sample items from sequential test 3.

Table 25

Summary of Analysis of Variance Comparing Adjacent Sequential Subtests

Source	SS	df	MS	F
Between Groups	62,861.42	4		
Comparison:				
1 with 2	73.35	1	73.35	21.51*
2 with 3	70.36	1	70.36	20.63*
3 with 4	71.86	1	71.86	21.07*
4 with 5	65.87	1	65.87	19.32*
Error (Within Groups)	187.58	55	3.41	
Total	63,049.00	59		

* $p < .001$

The results indicated that the difficulty levels of adjacent tests are significantly different from each other.

The sequence of administration of the sequential subtests is shown in Figure 2. Subtest 3 was administered first. Those students who scored above the criterion (six or seven items correct) on this subtest were routed to one of the more difficult subtests. The individuals who scored below criterion on subtest 3 were routed to one of the easier subtests. Finally, students whose scores met the criterion were assigned a hierarchical level of three and their testing was culminated. On the second day of testing, the remaining individuals were tested on either sequential subtest 1, 2, 4, or 5, depending on their score on subtest 3. Each subject who met the criterion on their assigned subtest stopped testing, while those subjects who surpassed or fell below the criterion were assigned another subtest on the last day of testing.

Confidence Test

Confidence testing represents a student response scheme for allowing more adequate evaluation of student knowledge than traditional testing methods. In confidence testing, a student can maximize his expected score by reflecting his degree of belief or probability that a specific response choice is correct. In the traditional procedure, using a

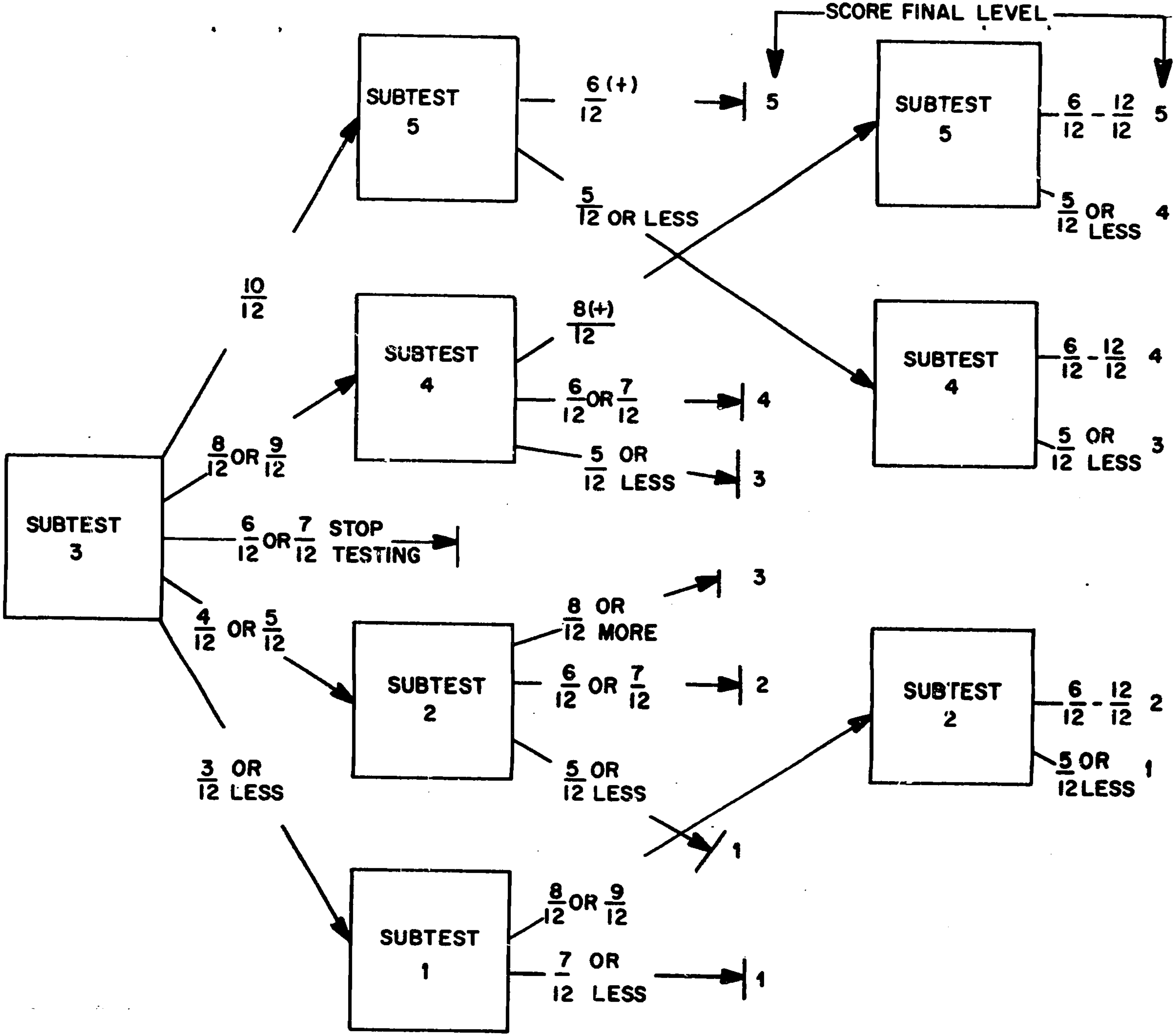


Figure 2. Sequential testing progression.

true-false test as an illustration, the student implicitly assigns a different probability for each response depending on his state of knowledge. If the student perceives the probability of "true" to be greater than .50, he should choose "true." If he perceives the probability to be less than .50, he can choose either response. Generally, a student with poor knowledge (p = .51) will get the same score (if correct) as the person with good knowledge (p = .90); therefore, the choice situation loses data about the student's knowledge.

The "pick one" method of confidence testing was used in the present investigation (Echternacht, Sellman, Boldt & Young, 1971). In the pick one method, the student first picks the answer he thinks is correct from among various multiple choice alternatives. He then assigns a probability value indicating his confidence. This probability value is then converted to an item score using special tables. (See AFHRL-TR 71-33 and AFHRL-TR-71-32.) In the tables, scoring is differentially weighted so that both item correctness and student confidence are reflected. Assume the student to assign a .7 probability to his answer. The tables indicate a score of .84 if his answer is correct and -.76 if his answer is wrong.

The main advantages of confidence testing lie in the more thorough assessment of student knowledge that it allows, in the virtual elimination of chance exerting an overriding effect on final test score, and in the increase in the student's perception of fairness that it provides. In addition, confidence scores can serve a training feedback purpose. If the situation arises in which the confidence scores of a group of students are very low, even though the absolute scores are high, an indication of the need for additional training would be evidenced.

The disadvantages of confidence testing are the scoring difficulty and increased time requirements for test scoring. However, those disadvantages are probably outweighed by the advantages. Moreover, computer scoring can be employed to ease the scoring burden.

As with the sequential test, the 30 items used in the confidence test were selected from the 200 items customarily used in the Block IV electronic principles tests. The items selected were representative of the total difficulty range (.40 - .99) and no item was included in both the sequential and confidence tests.

Characteristically, students have difficulty grasping the notion of assigning a confidence estimate to an answer. The formal test directions employed for the confidence testing are presented in Figure 3 and sample items are shown in Figure 4.

ELECTRONIC PRINCIPLES EXAMINATION

COMPONENT AND CIRCUIT CHARACTERISTICS

DIRECTIONS

This is a test of your knowledge of component and circuit characteristics. To answer each question, first, pick the answer you think is correct from among the multiple-choice alternatives. Then, assign a probability value indicating your confidence in that answer. The probability values you can use range from .25 to 1.00. If you assign a probability value of .25 to your answer, this indicates that you have selected randomly (guessed completely) and that you are only 25% sure that your answer is correct. Alternatively, if you assign a probability value of 1.00 to your answer, this indicates that you are 100% certain that your answer is right. If you assign a high probability value to an answer and the answer is correct, you will receive a higher score for that item than if you had assigned a low probability value for that answer (even though the answer is correct). On the other hand, if you assign a high probability value to an answer and the answer is wrong, a larger amount will be subtracted from your score than if you had assigned a low probability value to that wrong answer. In other words, you can get the greatest score by assigning high confidence estimates to those items you feel certain are correct and low confidence estimates to those items which you are uncertain about.

Examples

1. Which of the following is the best conductor of electricity?

   a. steel
   b. iron
   (c) copper
   d. rubber

   .90
   Confidence Level

2. Which of the following have a positive charge?

   (a) neutron
   b. electron
   c. neutrino
   d. proton

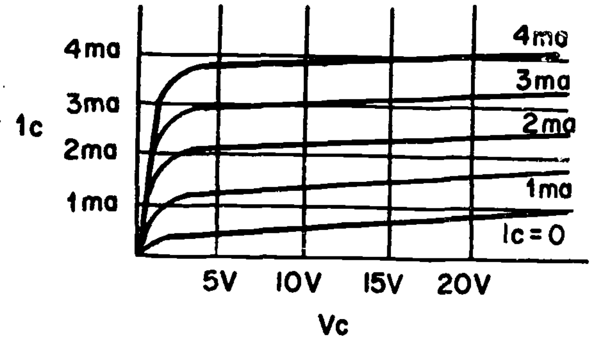
   .45
   Confidence Level

In the first example, the student was quite confident that his answer (c) was correct; therefore, he assigned a high probability value to that answer. In the second example, the student was somewhat uncertain of his choice; accordingly, he assigned a low confidence estimate to his answer (a). This was a good strategy, since the first answer was correct and the second answer was incorrect. If the student had reversed his confidence estimates, his score for the two items would have been much lower.

You will have 60 minutes to complete this test. The examiner will tell you when to begin.

Figure 3. Confidence test directions.

14. GIVEN:



COMPUTE ALPHA, USING V_C = 10 V.

a. .5

b. .9

c. 1.2

d. 1.6

____
Confidence Level

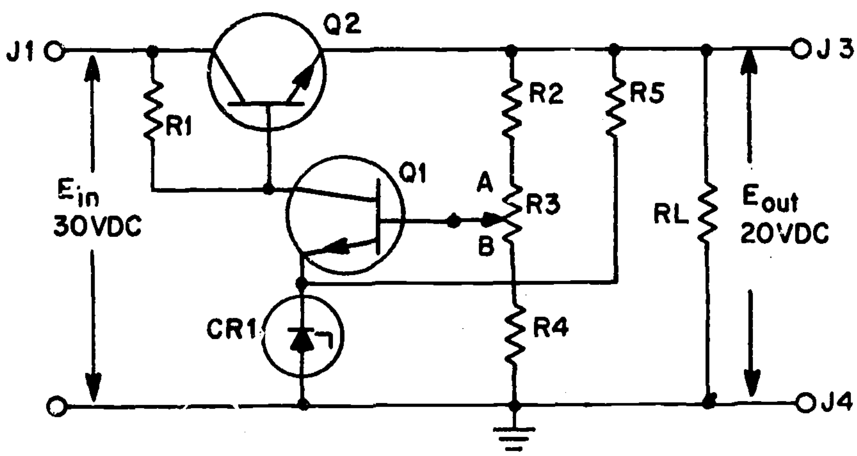
15. IN THE CIRCUIT SHOWN, THE OUTPUT VOLTAGE HAS DECREASED; THE CAUSE IS:

a. R4 open

b. Q1 open

c. CR1 open

d. R2 open



____
Confidence Level

Figure 4. Sample items from confidence test.

Pictorial Absurdities

In the electronic principles pictorial absurdities test, the student was required to examine nine circuit diagrams. In each diagram (Figure 5), some facet was depicted incorrectly. The examinee's task was to state what is wrong in the diagram.

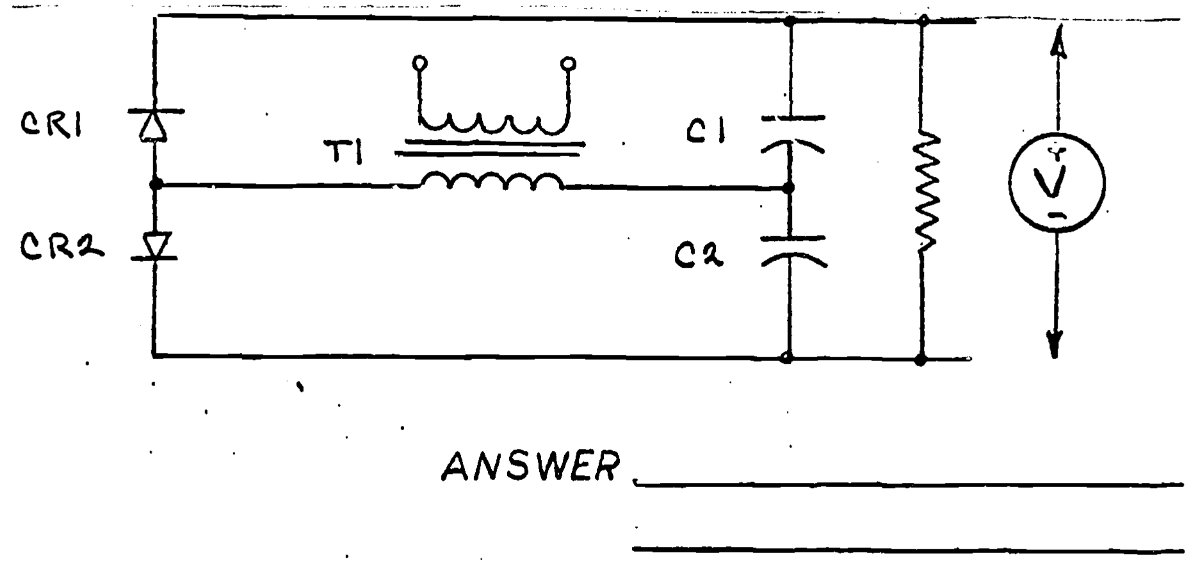


Figure 5. Sample item from pictorial absurdities test.

Absurdity items are particularly advantageous with persons who are handicapped in their ability to read. Additionally, absurdity items test the student's general level of alertness, his ability to recognize typical circuits, his knowledge of the circuit's characteristic function, his ability to concentrate, and his ability to separate important from irrelevant situational aspects. There is a difference in the electronic maturity of a student who answers correctly but gives a trivial answer as compared with the student who recognizes a major fault. Students who can recognize absurdities also are believed predisposed to the ability to troubleshoot electronic equipment adequately. This type of item possesses high face validity and, accordingly, is more acceptable than the usual multiple choice type of item in which language comprehension is heavily involved. The examinee instructions for the test are shown in Figure 6.

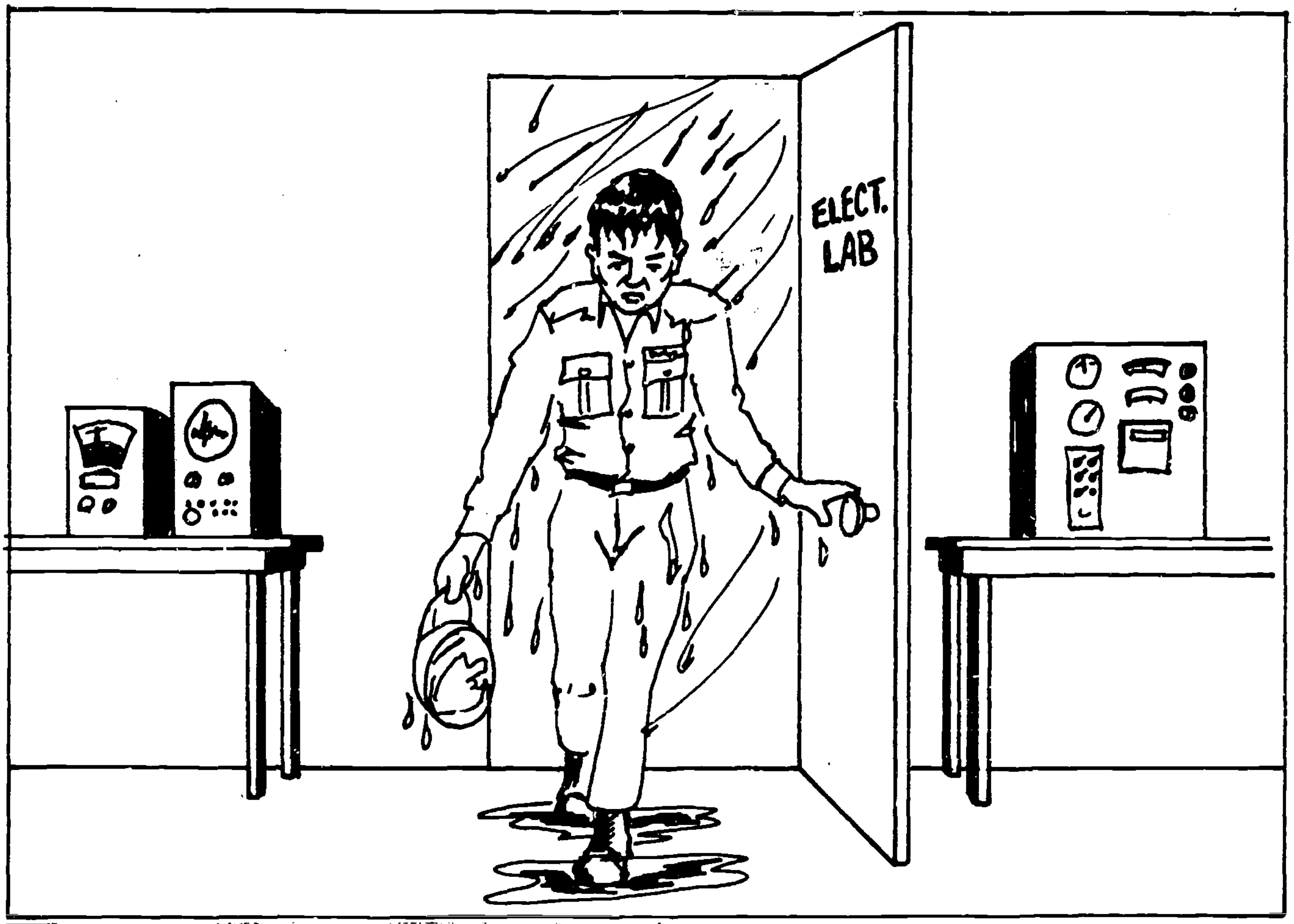
ELECTRONIC PRINCIPLES EXAMINATION

PICTORIAL ABSURDITIES

DIRECTIONS

This is a test of your ability to detect what is wrong with a picture or diagram. You will be presented with a series of nine pictures or diagrams and your task is to determine what is wrong with each. There is only one thing wrong in each picture. If more than one thing looks wrong, select the one which is most wrong.

Example



Answer ______________________________

The above example shows a man wearing wet clothes entering the lab. A correct answer for this picture is "Don't be wet when entering or working in lab."

You will be given two points credit for each correct answer. One point will be subtracted from your score for each incorrect answer. You will receive no points for unanswered questions. You will be allowed 15 minutes to complete this test. Remember, there is only one thing wrong in each picture.

Figure 6. Instructions for absurdities.

Pictorial Test (Schematic Reading/Component Identification)

To test the component identification factor through pictorial methods, a test of the examinee's ability to identify components in a diagram was employed. This type of item also possesses considerable congruency with the on-the-job aspects of the electronics technician's work.

Each item in the test consists of a diagram which is followed by one or more questions. First, the student examined each diagram. After he examined a diagram, he answered questions about the diagram. Fifteen minutes were allowed for completion of this test. Figure 7 presents a sample item from the pictorial schematic reading test.

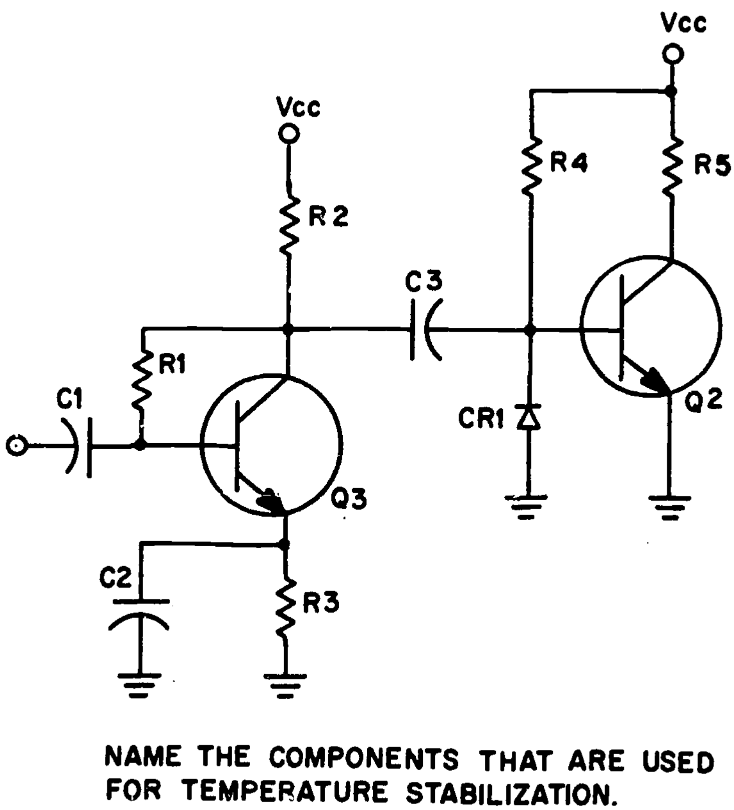


Figure 7. Sample item from the pictorial schematic reading test.

Cognition of Figural Systems

It seemed wise in any exploratory test development program to involve some pure factor approach to intellective function. The cognition of figural systems test is based on one aspect of the Guilford Structure-of-Intellect model. This model of intellective function has had extensive development and verification (Guilford & Hoepfner, 1971). The ability is loaded on the apprehension of spatial arrangements of items in one's psychological field. As employed here, the analogy might be drawn between the incidental development of "cognitive maps" (as suggested by Tolman in the maze learning situation) and the development on the part of the proficient electronics technician of "cognitive maps" of the equipment on which he works. The conjecture behind this test was that the better technician possesses a higher level of "visual insight," as the result of his exposure to equipment, than the poorer technician with the same experience.

In the cognition of figural systems test, the subject was presented with a series of component board layouts for electronic circuits. Each of the component board layouts had several circuit connections missing. Each circuit was verbally described at the top of each layout. The subject's task, then, was to draw connecting lines to form the indicated circuit. The student was allowed 30 minutes to complete this test. A sample item from the cognition of figural systems test, along with the correct response, is shown in Figures 8 and 9. The student directions for the cognition of figural systems test are shown in Figure 10.

The other advantages of this technique include its freedom from literacy requirements and its inherent interest to the examinee. In addition, the test involves a considerable deductive component.

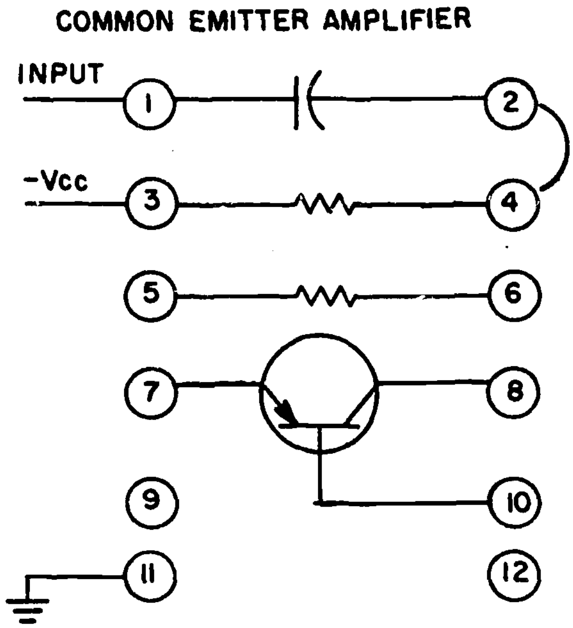


Figure 8. Sample item from the cognition of figural systems test.

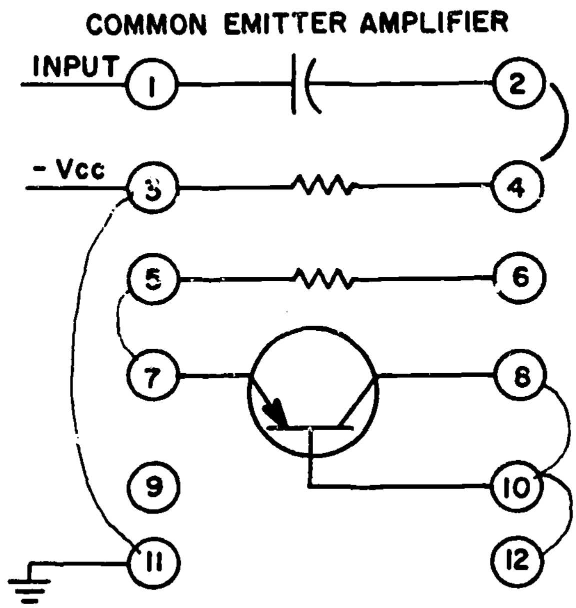


Figure 9. Correctly completed sample item from cognition of figural systems test.

ELECTRONIC PRINCIPLES EXAMINATION

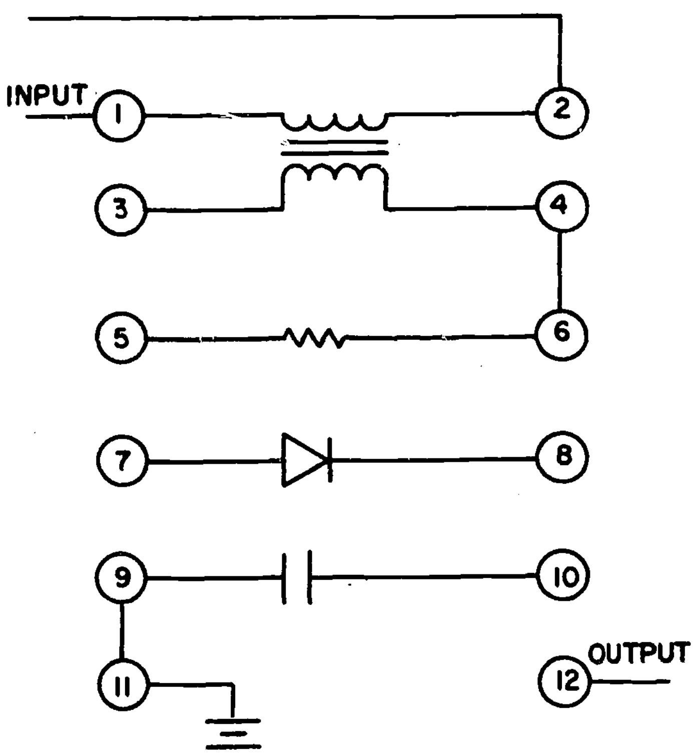
FIGURAL SYSTEMS

DIRECTIONS

This is a test of your ability to properly connect the various components and parts of a circuit. You will be presented with a series of component board layouts for electronic circuits. Several connections are missing in the layouts. The circuit is verbally described at the top of each layout. You are to complete the connections to form the indicated circuit. Use all components in the layout.

EXAMPLE

Half wave rectifier with capacitive filter



Complete the connections to form the indicated circuit in the above sample layout. After you have completed the circuit, turn to the next page where you will find the correct answer for this example.

Figure 10. Instructions for cognition of figural systems test.

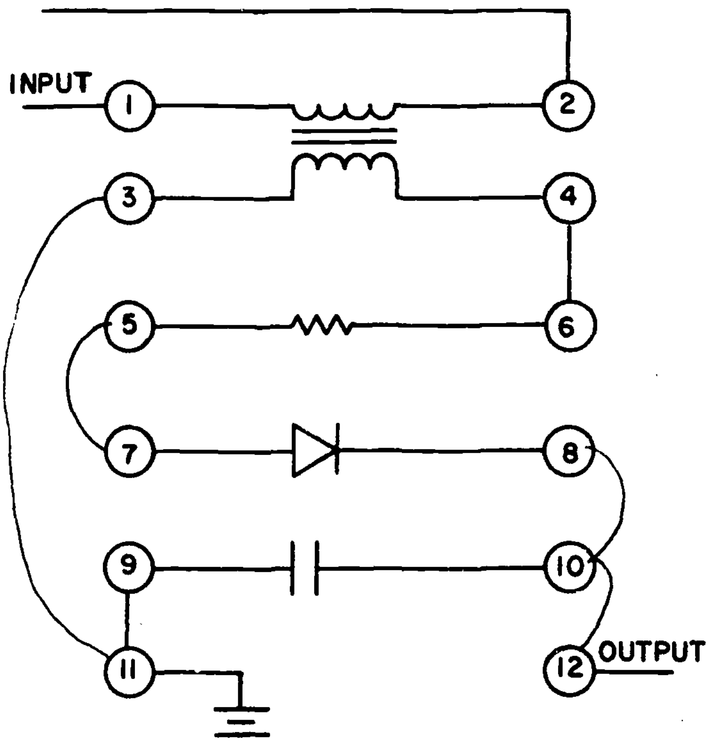


Figure 10. Instructions for cognition of figural systems test (cont.).

The student completing the example drew in the connections between 3-11, 5-7, and 8-10-12.

You will receive one point for each correct connection. One point will be subtracted from your score for each incorrect or missing connection.

You will be allowed 30 minutes to complete this test.

Electronic Analogies

The testing factor (factor 3) was also measured through an analogies test. In the electronic analogies test, the examinee was given a series of electronic analogies with one of the four members of the set missing. The examinee's task was to complete each analogy by selecting the correct item from a list of four items. The examinees were allowed 15 minutes to complete this test. A sample electronic analogies item is presented in Figure 11.

Analogies, as here employed, not only test knowledge (the examinee must recognize and understand the function and nuances of the circuits shown), but they also test the sophistication of his electronic thinking. To succeed on the analogies test, the examinee must be able to analyze circuits and to think logically. Deductive ability, also believed to be important to advanced electronic troubleshooting, is also involved in management of analogy items. Moreover, this type of test demands concentration and the ability to consider simultaneously several interrelated facts. These are also believed important to on-the-job function as an electronics technician. In addition, the electronic analogies test is unencumbered by literacy requirements.

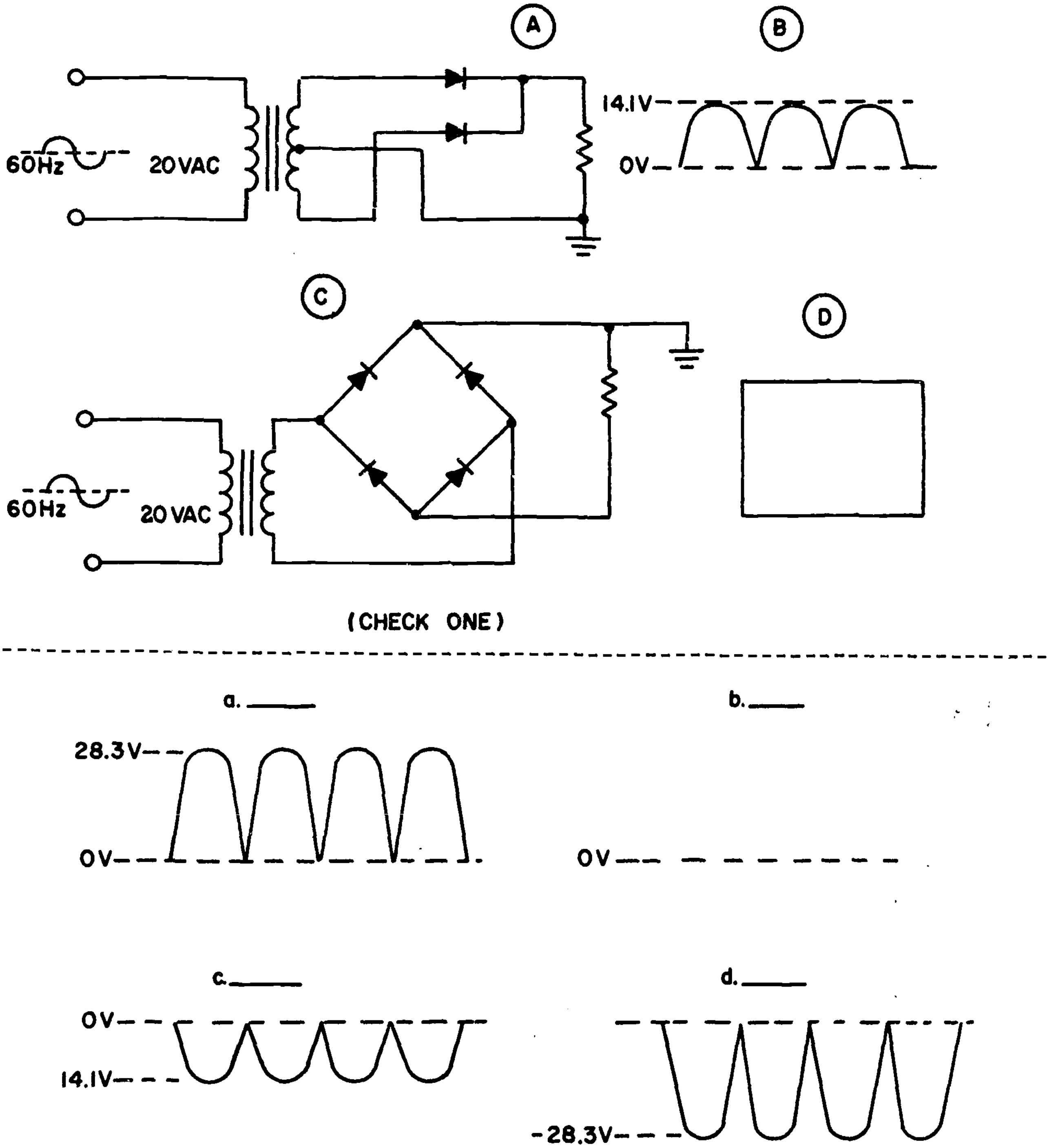


Figure 11. Sample electronic analogy items.

Signal Detection

Within the recent past and starting about 1954, the theory of signal detectability has emerged. Such theory possesses roots in both decision theory and in electrical engineering (with reference to ideal sensory devices). Swets and Tanner (e.g., Swets, Tanner, & Birdsall, 1964) generally (and, to our knowledge, first) have taken the lead in applying this approach to psychological problems. The novel aspect of theory of signal detectability, as employed by psychologists in reference to the human organism, is its emphasis on the sensitivity of the human, as separate from his role as a decision maker. Thus, persons concerned with theory of signal detectability suggest that a sensory threshold per se is a will-of-the-wisp and that we should really be concerned with response thresholds. The classical problem and approach, in the theory of signal detectability, are rooted in the detection of a signal in noise. The basic question is, " Assuming a knowledge of the conditional probabilities of the signal and the noise distributions, if the observer is presented with a stimulus, what is the 'best' rule for him to follow in deciding whether the stimulus is signal or noise? " The set of possible outcomes from such a trial is shown in Figure 12, where "yes" or "no" refer to the

	Response of Subject	
Stimulus Condition	Yes	No
signal	1	2
noise	3	4

Figure 12. Alternate outcomes from two choice experiment.

responses of the observer and where "signal" and "noise" refer to whether signal or noise was presented. If the observer says "yes" when a signal is presented (cell 1), we refer to his response as a "hit." If he replies "no" when a signal is present, we have a "miss" (cell 2 of Figure 12), and if he replies "yes" when noise is present (cell 3 of Figure 12), a "false alarm" situation is indicated. Finally, if the observer replies "no" when noise is present, a correct rejection is indicated. For a large number of goals (and ignoring payoff considerations, i.e., the cost of a "false alarm" or of a "miss" are the same as the gain from a "hit" or from a correct rejection), the best strategy for the observer to follow is governed by the likelihood ratio. The likelihood ratio is the ratio of the conditional probability of making a correct response given signal [$p(x|s)$] to the conditional probability of a correct response given noise:

$$\text{Likelihood ratio} = \frac{p(x|s)}{p(x|n)}$$

This ratio represents the best place for the observer to set his response threshold or criterion value. As the observer's criterion value deviates above or below the optimum, his "hit" or "false alarm" rate will vary.

If the "hit" and "miss" rate are plotted against each other, we have a receiver operating characteristic (ROC) curve. Such a curve is shown in Figure 13. Any observer whose response plot falls along the diagonal of the figure is responding randomly. The responses of the observer who responded "signal" too frequently would be represented by the y in Figure 13. This observer never misses a signal, but this success is at the cost of a high miss rate. The reverse holds for an observer whose responses fail at the point marked x in Figure 13.

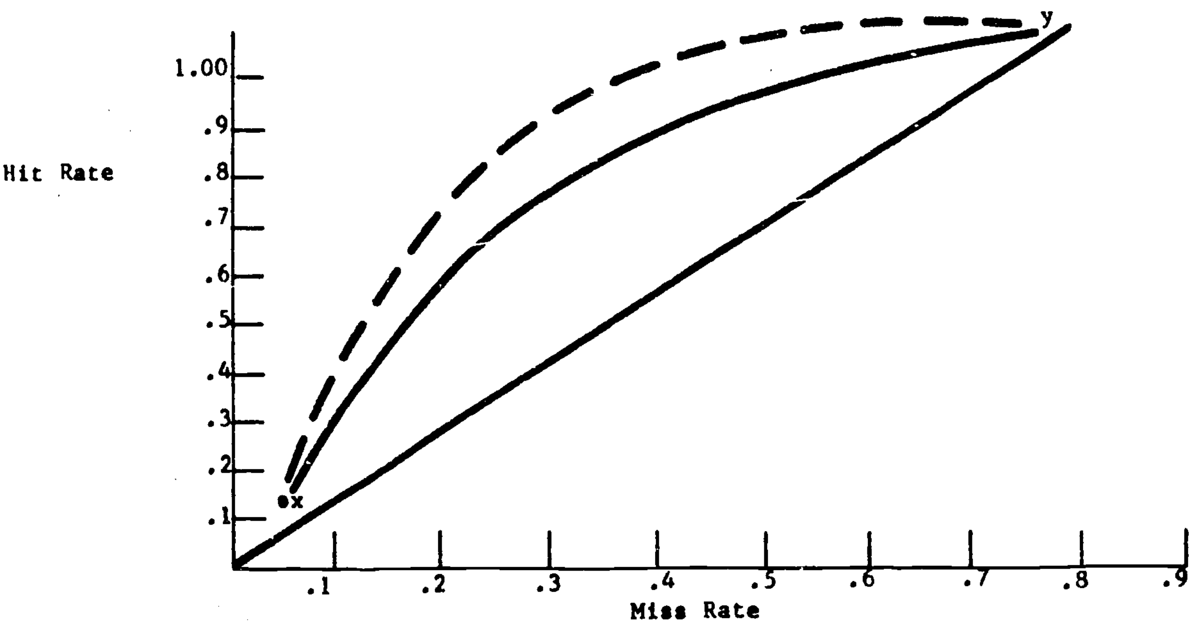


Figure 13. Receiver Operator Characteristic (ROC) curve.

Quite obviously, if the magnitude of the signal is increased and the noise distribution remains constant, the "hit" rate will increase, even if the criterion values are kept constant. The new curve will fall somewhere above the solid curve of Figure 13 and is indicated by the dashed curve. The observer's ability to react to the increased signal is an index of his sensitivity and is called, in theory of signal detectability, his d' value. Quite obviously, various observers will possess different d' values.

While the majority of signal detection experiments have used visual, auditory, or other sensory stimuli, the thinking behind signal detection theory, like threshold concept thinking, has recently been applied outside of the realm of sensory psychophysics. Some of these types of application have recently been summarized by Peterson and Beach (1967). Peterson and Beach cite examples for the signal detection theoretic approach for such areas as perceptual defense, recognition, memory, and judgment of the sources of short phrases from a magazine. According to these authors, "...these experiments show that it is possible to interpret a wide range of psychological phenomena within the framework of statistical decision theory."

Specifically, Applied Psychological Services (Siegel, Fischl, & Pfeiffer, 1968) has demonstrated the utility of d' and related variables as criteria for assessing performance and training effectiveness, as well as for predicting academic success in both a Navy and a college context.

Using responses to a true-false test situation (illustrated in Figure 14), these investigators demonstrated that it is possible to develop ROC curves which reflect differences in training effectiveness. Figure 15 presents the result for college students, and Figure 16 presents the results for military personnel.

In the preparation of Figures 15 and 16, the likelihood ratio (LX) was computed under the assumption of a normal probability density function for signal and noise curves and represents the ratio of the "hit" to "miss" probabilities at the point above which the observer placed his response criterion. The logic for this procedure has been discussed elsewhere by Swets, Tanner, and Birdsall (1964).

Using the "hit" and "miss" values, the detection sensitivity (d') was computed directly from the tables prepared by Patricia Elliott for the "yes-no" task, presented as Appendix 1 by Swets, Tanner, & Birdsall (1964). The higher the d' value, the greater is the distance between the means of the signal and the noise distributions and the less the overlap in distributions. Thus, high d' values represent a sensitive observer, one capable of distinguishing, with a low error rate, between signal and noise, i.e., one who has experienced effective training. Moreover, it was demonstrated that the differences noted among the d' scores at various levels of training cannot be attributed to differences among the subgroups in basic aptitude. Thus, the signal detection theory seems to possess considerable potential for evaluating training success.

Given Answer	Response of S		Total (f)
	True (A)	False (B)	
True (Signal)	f(S· A) ①	f(S· B) ②	f(True)
False (Noise)	f(N· A) ③	f(N· B) ④	f(False)
Total (f)	True	False	

Definitions:

$$P(A|S) = \frac{f(S\cdot A)}{f(\text{True})} = \text{Hit rate}$$

$$P(A|N) = \frac{f(N\cdot A)}{f(\text{False})} = \text{False alarm rate}$$

Figure 14. "True-false" response classification matrix.

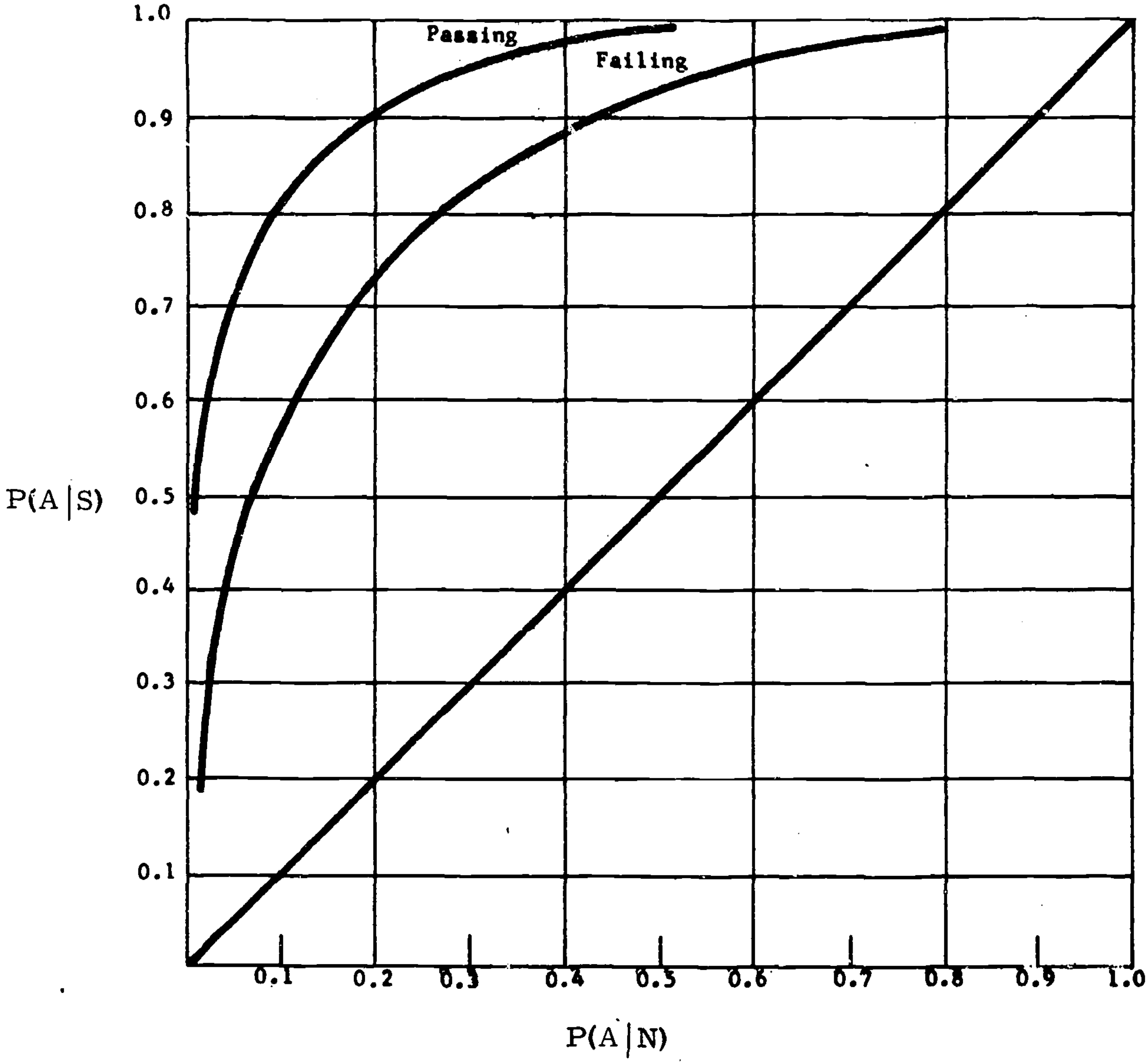


Figure 15. ROC curves of 20 passing and 20 failing college students.

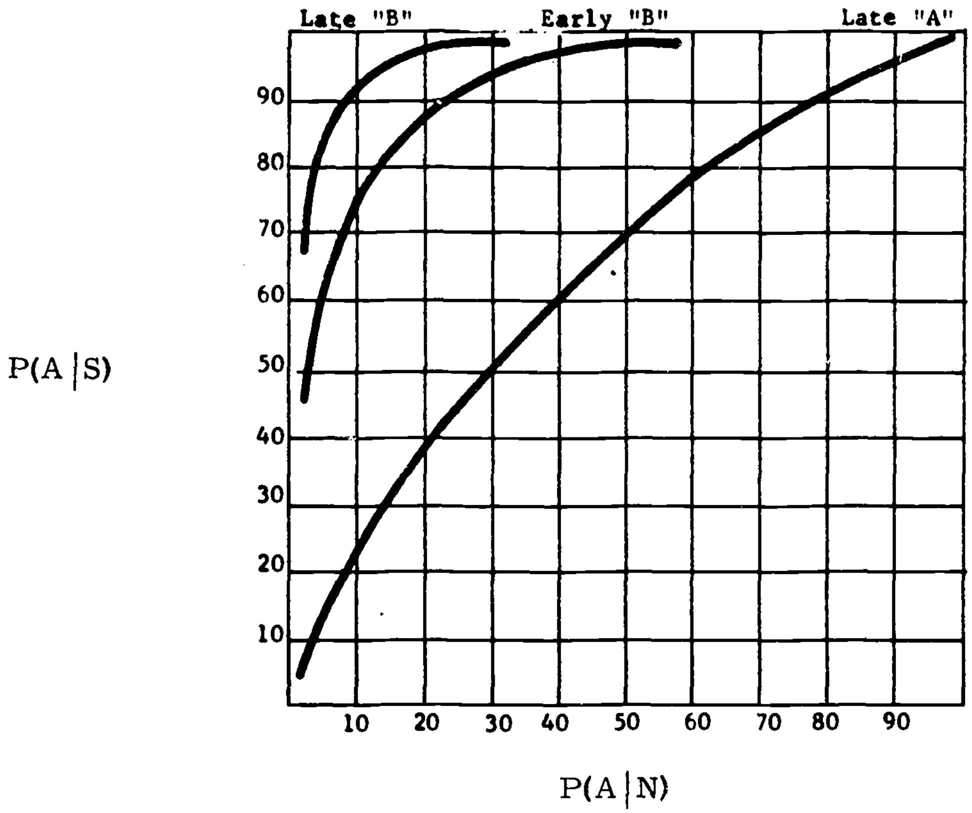


Figure 16. ROC curves for electronics maintenance personnel at three levels of training experience.

In the present investigation, a multiple choice confidence test was administered to the electronics students. This test was scored in both the conventional manner and in accordance with theory of signal detection principles.

Administrative Specialist Tests

The advanced measurement techniques developed for measuring the administrative factors were: (a) partial knowledge, (b) technical words, (c) Thurstone scaled typing samples, (d) pictorial absurdities, (e) signal detection, (f) need achievement, and (g) psychophysiological arousal. The development and use of each of these testing techniques are described in detail in paragraphs which follow. All of these tests, except need achievement and psychophysiological arousal were oriented to test the document preparation factor (factor 1).

Partial Knowledge Testing

Traditionally, in scoring a four-choice multiple choice question, a student is given credit for the correct answer and no credit for a choice of any incorrect answer or distractor. Partial knowledge exists when the student can identify one or more of the distractors. With the partial knowledge scoring technique in a multiple choice format, one point is given for each distractor identified and three points subtracted if the correct answer is identified as a distractor. Scores on each four-choice item can range from plus three to minus three. Partial knowledge testing, as in confidence testing, yields increased item and test variance and penalizes for random guessing. Moreover, as compared with the usual test scoring procedures, it provides a greater source of feedback information in regard to areas of student misunderstanding of subject matter content.

The 25 items used in the partial knowledge test sampled the range of knowledge required of students in Block II of the course for administrative specialists. Twenty minutes were allotted for this test.

As with confidence testing, students sometimes experience difficulty learning the idea of selecting the incorrect distractors, rather than choosing the correct answer as in the multiple choice format. The burden of explanation, then, is placed on the test directions and the administrator. Figure 17 presents the examinee directions employed for the administrative specialist partial knowledge test. Sample items taken from the partial knowledge test are shown in Figure 18.

DIRECTIONS

This is a test of your knowledge of document preparation. Your task is to identify as many of the incorrect answers from among the multiple choice alternatives as you can. If you identify all three incorrect alternatives, you will get a higher score than if you only identify two of the incorrect alternatives. Identifying two incorrect alternatives gives you more points than if you identify only one incorrect alternative. Accordingly, for each incorrect alternative you identify, you will be awarded one point. Conversely, if you identify the correct alternative as being incorrect, then, three points will be subtracted from your score. For example, if you correctly identify two of the wrong answers in a four choice multiple choice question you receive two points. If, on the other hand, you correctly identify two of the wrong answers and incorrectly identify the correct answer, then you receive a score of -1 as shown below:

$$1 + 1 - 3 = -1$$

If you identify all three wrong answers your score is +3 as shown below:

$$1 + 1 + 1 = 3$$

Correct identification of all three wrong answers produces the highest possible score on a question. In order to identify the wrong answers, simply place a checkmark next to each answer youfeel is incorrect.

Example

1. Identify the incorrectly spelled word (check the wrong answers).

   a. _____ typwriter
   b. _____ typewriter
   c. __✓__ tipewriter
   d. __✓__ tipwriter

In the above example, the student was quite certain that choices c and d were incorrect; therefore he obtained two points for that question. If he had chosen choice b in addition to choices c and d, his score would have been -1, since he incorrectly identified the correct answer as being incorrect. Your best strategy, then, is to check only those answers you are certain are incorrect.

You will have 20 minutes to complete this test. Remember, only select the incorrect alternatives, or the alternatives that are most incorrect. The examiner will tell you when to begin.

Figure 17. Partial knowledge test directions.

13. How many net words per minute did you type if you typed 183 strokes, in five minutes, with five errors? (check the incorrect alternatives)

a. _____ 5/5
b. _____ 7/5
c. _____ -3/5
d. _____ -13/5

16. On air force letters there is a: (check the incorrect alternatives)

a. _____ one inch left margin
b. _____ one-half inch left margin
c. _____ inch and a half left margin
d. _____ two inch left margin

Figure 18. Sample items from the partial knowledge test.

Technical Words

Vocabulary in a content area has been a traditional reflector of sophistication in the area. Technical words tests generally correlate most highly with other measures of achievement. Just as a man's general vocabulary reflects his general intelligence, his technical vocabulary reflects his level of accomplishment in a technical area.

The 32 items in the technical words test sampled the total range of knowledge required of students in Block II of the administrative specialist course.

In the technical words test, the examinee was presented with a word, phrase, or symbol followed by four multiple choice alternatives. His task was to identify the multiple choice alternative which is synonymous with or gives a major use for the word, phrase, or symbol. Fifteen minutes were allotted for completion. Figure 19 presents some sample technical words items.

5. Ampersand (check one)

a. _____ +
b. _____ &
c. _____ #
d. _____ %

8. Shift key (check one)

a. _____ is used to change lines
b. _____ is used to space
c. _____ is used to capitalize
d. _____ is used to indent

Figure 19. Sample technical words items.

Thurstone Scale of Typing Ability

The Thurstone scaling concept, which produces an equal appearing interval scale along a continuum, has been extensively employed in investigations into attitudes but has been seldomly applied for student measurement purposes. Yet, the production and use of such a scale along a subjective difficulty continuum possesses attractive possibilities as applied to typing achievement measurement. Such a scale would be rooted to the difficulty of the items as perceived by the instructors. Accordingly, it can be argued that a student who achieves a given score on a test at level 4 along a Thurstone scaled set of typing materials can be considered, in the view of the instructors, to be twice as proficient as the student who receives a comparable score at level 2. Moreover, the scaling properties make the conception attractive from the point of view of test economy and student placement. From the test economy point of view, the student can be moved along the scale in equal difficulty increments to find his maximum level in minimum time. From the student placement point of view, it affords a direct approach to homogeneous grouping.

To develop a set of Thurstone scaled typing items, a set of 18 typing samples of varying difficulty were collected from several sources at Keesler AFB. These typing samples were then submitted to 12 typing instructors, at the Administrative Specialist School, who were asked to judge the difficulty of each sample along an 11-point scale. The formal directions are shown in Figure 20.

The median difficulty value of each typing sample on the 11-point scale was then determined, along with the item's interquartile range.

In the Thurstone scaling approach, the median is considered to be the scale value of that stimulus. The interquartile range is considered to be a measure of uncertainty. Agreement among the judges on the scale value produces a low interquartile range value, whereas disagreement produces a large value. Figure 21 shows the scale and interquartile range values of each typing sample.

Name____ Today's Date__________

DIFFICULTY DETERMINATION

You have been presented with a series of typing samples. These typing samples vary in difficulty from very easy to very difficult. Your task is to assign a difficulty level to each typing sample. Down the left hand side of the page is a numerical listing of each typing sample. Across from each typing sample number is an eleven point scale. You are to place a checkmark (✓) at that point on the scale that represents the difficulty level of the typing sample. Try to vary your checkmarks so that some appear in all eleven columns. Do not hesitate to use extreme responses numbered 1 and 11, if you feel any typing sample deserves one of them.

EXAMPLE

	Very Easy		Moderately Easy				Moderately Difficult			Very Difficult	
	1	2	3	4	5	6	7	8	9	10	11
Sample 1	__	__	__	✓	__	__	__	__	__	__	__
Sample 2	__	__	__	__	__	__	__	__	__	__	✓

The first check indicates that the judge assigned a Moderately Easy difficulty level to typing Sample 1. The second check indicates that the judge thought typing Sample 2 was Very Difficult.

Figure 20. Direction sheet for Thurstone scaling of typing samples.

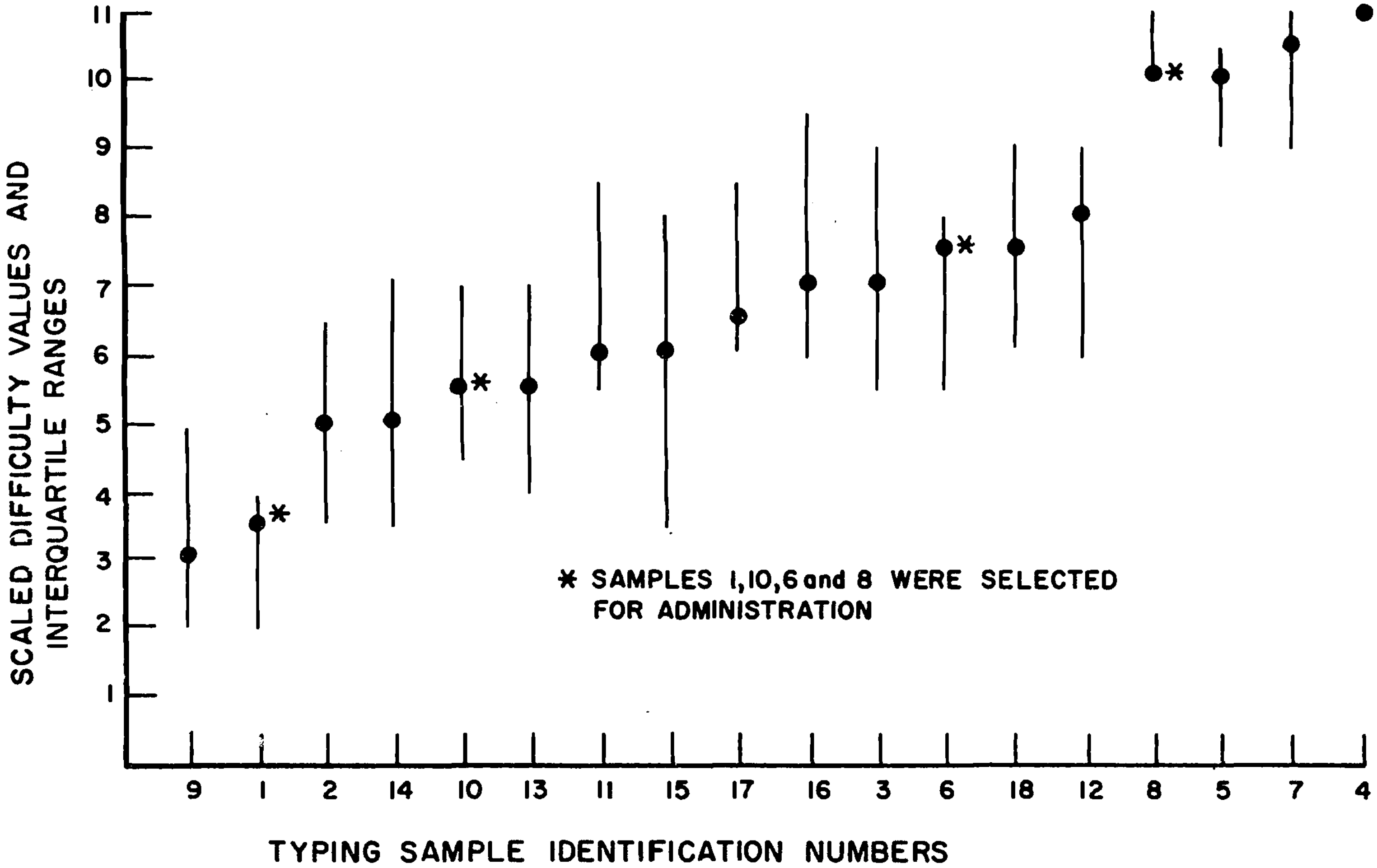


Figure 21. Interquartile ranges and median scale values for 18 typing samples judged for difficulty by 12 typing instructors.

Samples 1, 10, 6, and 8 were selected for inclusion into the administrative specialist test battery. These typing samples have scale values of 3.5, 5.5, 7.5, and 10.0 with interquartile range values of 2.0, 2.5, 2.5, and 1.0 respectively. We note that (see Figure 21) the upper bound of the interquartile range value for any selected sample does not overlap with the lower bound of the next selected sample.

Typing sample number 6 which had a difficulty value of 7.5 is shown in Figure 22. Ninety seconds were allowed for completion of this typing sample. The typing samples were administered to the students in order of difficulty from easiest to most difficult. Rather stringent time limits were set to ensure that no examinee would complete any typing sample before the allotted time limit. Otherwise, faster students would be penalized.

....:....10....:....20....:....30....:....40....:....50....:....60....:....70	Stroke Count
On the morning of 24 April the Germans counter attacked with a	69
force estimated at 2 battalions of infantry supported by artillery.	138
The attack lasted from 0800 to 1200, during which time the enemy made	208
assaults from practically every direction. In this action the 2d	274
Battalion lost 21 dead and 11 wounded. The gallant and intrepid	340
conduct of this entire battalion afforded a great tactical advantage	409
in seizing terrain and in the defeat of German arms in North Africa.	478
....:....10....:....20....:....30....:....40....:....50....:....60....:....70	

Figure 22. Thurstone scaled student typing sample.

Pictorial Absurdities

In the administrative absurdities test, the student was required to examine 10 pictures or documents. Each contained some incorrect action or facet. The examinee's task was to state what is wrong in each picture. Fifteen minutes were allotted for this test. The advantages of this type of test were discussed under the prior description of the electronic principles tests. A sample item from this test is shown in Figure 23.

Signal Detection

With regard to the Administrative Specialist tests, the partial knowledge test was scored using signal detection principles. The theory of signal detection was discussed, at length, in the descriptions of the electronic principles advanced measurement techniques.

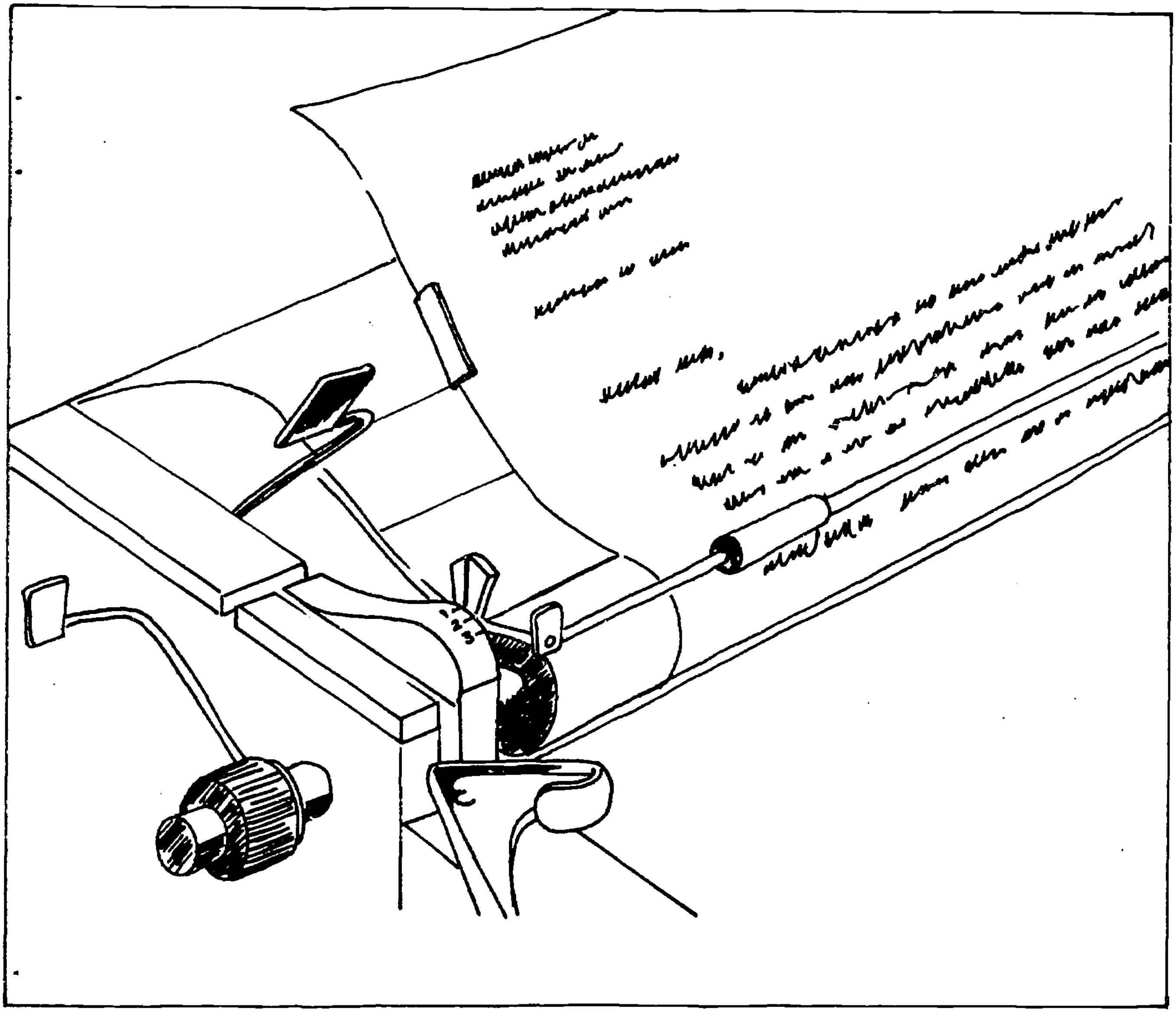


Figure 23. Sample administrative absurdities item.

Auxiliary Assessment Methods

The remaining two measures used in the current investigation were not tests, per se, but are novel measures which might predict or moderate future performance or which might provide training feedback information beyond that usually available. These two additional measures involve the concepts of need achievement and psychophysiological arousal.

Need Achievement

An airman's motivation, whether in the school situation or in the field, can play an important a role as a success determinant. Accordingly, it seemed important to explore whether or not a measure of motivation could make a significant contribution to the goals of the present study. The measure of motivation employed was the desire or drive of the student to attain a prestigious position in general and in the Air Force.

Atkinson and O'Connor (1963) correlated the prestige value of occupations with probability of success estimates given by high school students. The resultant correlation was found to be .90. In a similar study, Strodtbeck, McDonald, and Rosen (1957) found that Jewish boys would be more satisfied with high status occupations (e.g., doctor, executive) and less satisfied with low status occupations (e.g., night watchman) than Italian boys. In other words, a steeper slope of satisfaction was found for Jewish boys than for Italian boys. These authors also found a steeper slope of satisfaction for middle class boys than for lower class boys. Atkinson (1966) feels that the Strodtbeck, McDonald, and Rosen study is "... particularly exciting, because independent studies have shown n Achievement to be stronger among Jews than Italians and stronger in the middle class than in the working class (or upper class)" [p. 164].

A short "need achievement" questionnaire was developed and administered to both the electronics and the administrative samples. The data yielded by this questionnaire allow derivation of insight into the relationships between student motivation to succeed and test scores. The questionnaire used employed the methods and techniques described by Morgan (1966) which were based upon the Strodtbeck, McDonald, and Rosen (1957) study.

In the questionnaire, the students were required to compare the desirability of various civilian and Air Force occupations on a six point scale. Then, the regression or slope of the individual on the group is calculated. The resultant slope index is the index of achievement motivation. Accordingly, the greater the covariance between the individual's occupational hierarchy and that of the group, the higher his achievement motivation index. The civilian occupations questionnaire is shown in Figure 24.

OCCUPATIONS QUESTIONNAIRE

Name (print) ____________________ Date ____________________

Directions

We are interested in how people compare occupations. How do you think most people would feel if a friend or relative of theirs chose certain types of work? Decide whether most people think each of the listed occupations is: very desirable (VD), desirable (D), slightly desirable (SD), slightly undesirable (SU), undesirable (U), or very undesirable (VU).

Civilian Occupations
(Circle one for each occupation)

a. Night Watchman	VD	D	SD	SU	U	VU
b. Carpenter	VD	D	SD	SU	U	VU
c. Mail Carrier	VD	D	SD	SU	U	VU
d. Bus Driver	VD	D	SD	SU	U	VU
e. Bookkeeper	VD	D	SD	SU	U	VU
f. Druggest	VD	D	SD	SU	U	VU
g. Teacher	VD	D	SD	SU	U	VU
h. Doctor	VD	D	SD	SU	U	VU
i. Assemblyline Worker	VD	D	SD	SU	U	VU
j. Janitor	VD	D	SD	SU	U	VU
k. Office Manager	VD	D	SD	SU	U	VU

Figure 24. Civilian occupations need achievement questionnaire.

Psychophysiological Arousal

The psychophysiological measure used in this investigation was the proportion of active sweat glands on a four millimeter square area of the central whorl on the finger tip of the third finger of the less used hand. The technique was developed by Sutarman and Thompson (1951) as a substitute for the galvanic skin response (GSR) and has since been improved by Johnson and Dobbs (1967). Only one investigator (Martens, 1969) has successfully used it as a test of social facilitation theory. Another investigator (Bergman, 1971) used it to assess the physiological response to crowding and test failure.

To employ the technique, a solution is applied to the finger tip of the subject with a cork stopper. After a 15 or 20 second interval, the solution dries and is removed with Scotch Magic Transparent Tape and mounted on millimeter ruled graph paper. When the print is placed against a lighted surface and subjected to 10-20 power magnification, the number of active and inactive sweat glands on a demarcated area of the sweat print can be counted. The proportion of active sweat glands was used as the criteria rather than the number of active sweat glands since the total number of sweat glands on a finger tip can vary from individual to individual. The chemicals used to make the sweat sensitive compound were as follows:

1. five grams of polyvinyl formal which withdraws from moisture

2. ten milliliters of butyl phthalate which gives the print toughness

3. twenty grams of a semi-colloidal dispersion of graphite in trichloethylene which provides optical contrast

4. one-hundred milliliters of ethylene dichloride which is a solvent

The interscorer reliability coefficients for the number of active sweat glands (Johnson & Dobbs, 1967) were reported to range from .87 to .98, and the reliability of one scorer who made counts several days apart was .99.

Johnson and Dobbs (1967) report increases in palmar sweating before and after the administration of mathematics tests to students and a decrease in sweating during the tests. Harrison and MacKinnon (1962, 1966) found that palmar sweating decreased while an individual suspended an unsupported limb in the air, and that the anhydrous response was mediated by the adrenal medulla.

Since many persons exhibit emotional arousal in testing situations, it was thought relevant to include a measure of psychophysiological arousal in this investigation. As with need achievement, arousal may account for some portion of the test score variance. On difficult or unfamiliar tasks, the performance of highly aroused subjects should be disrupted, but moderate arousal should be facilitating. Similarly, no arousal at all should exert a negative effect on performance. Application of the solution is quick and easy, and the technique is much less obtrusive than other arousal measures (e.g., GSR) [Bergman, 1971].

Questionnaires and Interviews

Following completion of the testing, each student was required to complete a structured questionnaire regarding the desirability or undesirability of each of the advanced measurement techniques employed. These questionnaires are reproduced in the appendix to this report.

Additionally, instructors from the electronics school and instructors from the administrative school were given individual briefings on the development, rationale, and use of each advanced measurement technique. At the completion of the briefings, 10 instructors from each school were individually interviewed regarding the probable effectiveness of the new techniques as compared with the old, the utility of the novel techniques for assessing student attainment of training objectives, administrative and logistic considerations, and possible effects of the new techniques on the role of the instructor and the training evaluation specialist.

Data Treatments and Criteria Employed in Judging the Value of the New Instruments

Correlational Analysis

To test the merit of the novel test instruments, each of the advanced measurement techniques and the auxiliary measures was considered as a predictor within a concurrent validational effort. The criteria in this validation effort were final average and examination grades in the respective courses. The advanced measurement techniques were anticipated to account for a significant proportion of the final average and examination grade variance. As an ultimate test of the merit of the novel measurement techniques, predictive validity would have to be established. However, we note that such predictive validity has not been established for the usual course examinations. Moreover, it seems that because the novel instruments tap a greater range of cognitive function, on the surface they should possess greater predictive validity than the currently employed measures. It seems obvious that posttraining field performance involves a greater range of intellective activity than is measured by the usual multiple choice tests.

Reliability

The second test of the merit of the novel measures was concerned with test reliability. Where appropriate, reliability estimates based on the Kuder-Richardson (formula 20) (split half methods were employed in those instances in which the Kuder-Richardson formula could not be utilized) and Spearman-Brown methods were applied to each advanced measurement technique.

Cost/Benefit Analysis

The third test of the value of the present advanced measurement instruments involved a cost benefit analysis. The cost/benefit analysis relied on the derivation of a "figure of merit" on the basis of the following formula:

$$\text{Cost/Benefit} = \left[\frac{DC + AC + SC}{N}\right] \times (1-r^2) \text{ or } \frac{1}{S}$$

where: DC = Direct costs
AC = Administrative costs
SC = Support costs
N = Number of students evaluated
r^2 = Test-criterion variance
S = Number of course improvement suggestions elicited

This formulation was applied to each test type separately by course.

III RESULTS

The first set of results to be presented are those dealing with the electronic measures. These results will be followed by a presentation of the administrative specialist results. Each set of results is composed of five parts, including: (a) correlational analysis (concurrent validity), (b) reliability, (c) student questionnaire results, and (d) cost/benefit analysis.

Electronic Principles Measures

Electronic Principles Correlational Analysis

Each of the advanced measurement techniques was used as a predictor of school performance in a concurrent validity study. The auxiliary measures (stress measure and need achievement measure) and the scores on the usual school Block IV multiple choice tests were also included as predictors in this phase of the analysis. The criterion of school performance was the average of the final grades for electronic principles in Blocks I-VII excluding Block IV. (Only the first seven blocks of instruction were considered, inasmuch as these were the only blocks that all the students had in common.) The Block IV exam was not included as part of the criterion average since it was included as a predictor. This validational paradigm is analogous to item analytic techniques in which each item is correlated with total test score after the item's contribution to that test score has been removed.

As a second step in the correlational analysis, the method of stepwise regression (Dixon, 1965) was employed. Stepwise regression computes a series of "... multiple linear regression equations.... At each step one variable is added to the regression equation. The variable added is the one which makes the greatest reduction in the error sum of squares. Equivalently, it is the variable which has highest partial correlation with the dependent variable partialled on the variables which have already been added; and equivalently it is the variable which, if it were added, would have the highest F value" [p. 233].

Prior to completion of the stepwise regression analysis, several of the advanced measurement technique scores were either multiplied by a constant and/or had a constant added to them. This was performed in order to eliminate decimals and negative numbers. These measurement techniques with their additive and/or multiplicative constants are shown in Table 26.

Table 26

Additive and/or Multiplicative Constants for the Electronic Principles Advanced Measurement Techniques

Technique	Additive Constant	Multiplicative Constant
Figural Systems	101.00	
Schematic Reading		10.00
Confidence Score	134.00	10.00
Confidence Test (d')	1.55	10.00
Analogies	8.00	10.00
Achievement Motivation		10.00

The means and standard deviations of the electronic advanced measurement techniques, the auxiliary measures, the Air Force aptitude variables, the Block IV score, and the criterion variables are shown in Table 27. The intercorrelation matrix for these same variables is presented in Table 28.

From the point of view of the concurrent validity of the advanced measurement techniques, we note that, with the exception of the absurdity tests (where no validity was shown), moderate to rather high coefficients are indicated in Table 28. The conventional score, confidence score, and d' measures exhibited particularly high validity coefficients--.64,.58, and .64 respectively. The remaining advanced measurement techniques exhibited validity coefficients which ranged from .28 (analogies) to .42 sequential test level. These are interpreted as moderate and acceptable. The auxiliary measures, need achievement and psychophysiological arousal, exhibited no predictive validity whatsoever.

Table 27

Mean and Standard Deviation of the Advanced Measurement Techniques, the Auxiliary Measures, the Air Force Aptitude Variables, the Block IV Score, and the Criterion Variables for 41 Students Enrolled in the Electronics Principles Course

Variable	Mean	Standard Deviation
Sequential Test Level	2.829	1.093
Confidence Test Score	159.073	65.999
Signal Detection Test (d') Score	196.585	67.697
Conventional Test Score	15.439	5.201
Pictorial Absurdities	3.268	2.398
Pictorial Schematic Reading	85.341	22.742
Cognition of Figural Systems	48.244	17.121
Analogies	43.049	19.497
Need Achievement (Civilian)	144.220	84.641
Need Achievement (Air Force)	132.561	28.584
Psychophysiological Arousal	73.366	26.261
General Aptitude	78.780	12.134
Mechanical Aptitude	77.073	12.989
Administrative Aptitude	71.098	17.906
Electronic Aptitude	86.341	5.365
Block IV Exam Score *	77.512	12.789
Criterion	79.463	7.349

* Solid State Power Supplies and Amplifiers

Table 28

Intercorrelation Matrix* of the Advanced Measurement Techniques, the Auxiliary Measures, the Air Force Aptitude Variables, the Block IV Score, and the Criterion Variables for 41 Students Enrolled in the Electronics Principles Course

Variables	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Sequential Test Level	463	519	524	-125	123	300	120	146	037	-013	-007	-080	074	083	223	418
2. Confidence Test Score		946	885	-044	121	277	250	032	-039	074	320	360	073	172	414	574
3. Signal Detection Score			884	-020	180	352	246	026	018	001	314	378	074	230	421	635
4. Conventional Test Score				-044	145	303	227	-132	-041	064	227	336	-048	319	487	636
5. Pictorial Absurdities					-127	-005	-303	-007	140	094	102	058	-170	107	296	-031
6. Pictorial Schematic Reading						126	119	111	128	076	188	-033	143	234	-044	294
7. Cognition Figural Systems							419	008	218	081	121	401	-127	384	337	382
8. Analogies								133	156	-111	345	170	190	387	135	279
9. Need Achievement (Civilian)									-152	-442	-013	-244	157	-270	-221	-228
10. Need Achievement (Air Force)										129	120	-134	199	116	183	083
11. Psychophysiological Arousal											071	201	-094	125	235	056
12. General Aptitude												405	461	515	175	329
13. Mechanical Aptitude													-104	443	280	350
14. Administrative Aptitude														192	013	108
15. Electronics Aptitude															301	399
16. Block IV Exam. Score																486
17. Criterion																

* Decimal points have been eliminated.

In order to bring organizational clarity to the data of Table 28, the table was cluster analyzed. One of the advantages of a cluster analysis, in this context, is that it can demonstrate why some of the variables which are highly correlated with the criterion are either not present in the multiple regression equation or contribute only a minimal amount of variance to multiple prediction of the criterion score. The results of the cluster analysis are shown in Table 29.

The largest most highly interrelated cluster consists of four methods based on items from the Block IV tests. One would expect, then, a great deal of variance overlap between these tests. The fourth measure (sequential test) in this cluster consists of different Block IV items from the same tests. Surprisingly, the usual Block IV exam score failed to fall into this cluster.

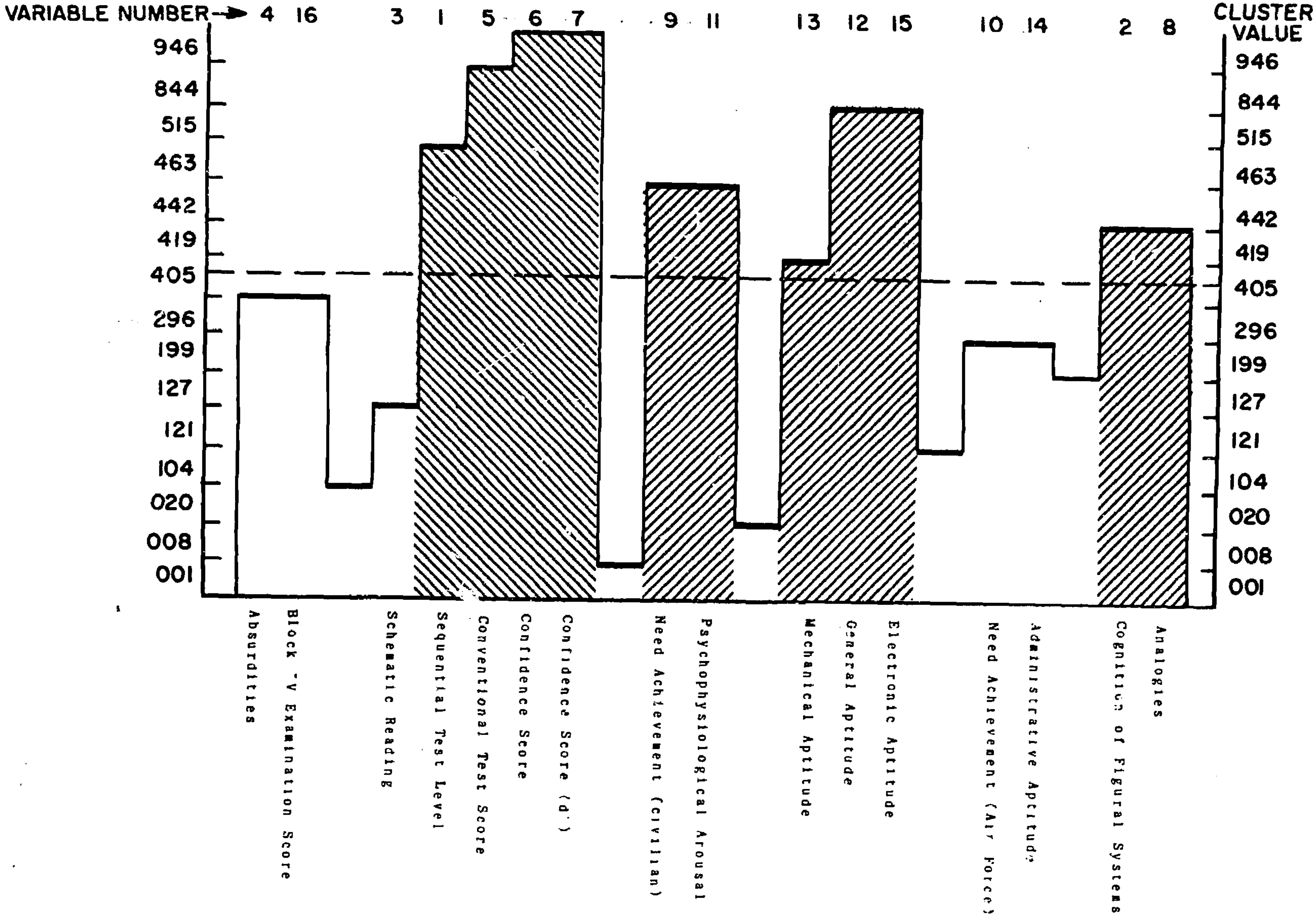
The next largest cluster consists of the electronic, general, and mechanical aptitude tests from the Air Force aptitude battery.

The third cluster consists of the psychophysiological arousal measure and the civilian need achievement inventory. A low score on the psychophysiological measure indicates high arousal while the need achievement score was polarized in the opposite direction. Those two variables are inversely related to each other at a value of -.44. This finding suggests, as might have been anticipated, that airmen who exhibit considerable physiological arousal are also high in achievement motivation. Conversely, airmen who exhibit little physiological arousal during test taking are low in achievement motivation. These two tests exhibited insignificant, but consistent, relationships with the mechanical aptitude, the electronics aptitude, and the Block IV test scores. The need achievement index was correlated -.24, -.27, and -.22 with each of these three tests respectively, while psychophysiological arousal correlates .20, .13, and .24 with the three tests. The above results seem to be consistent with current motivational theory. Apparently, a certain degree of arousal is necessary for motivated behavior.

The final cluster consists of the figural systems and analogies tests. Both of these tests require a high degree of abstract thinking by the examinee. What is unusual about both of these tests is that they contain items which are almost devoid of verbal content. Although these tests are nonverbal, they still correlated with the criterion .38 and .28, respectively.

TABLE 29

CLUSTER ANALYSIS OF 16 PREDICTOR VARIABLES DERIVED FROM THE SCORES OF 41 ELECTRONICS PRINCIPLES STUDENTS



The remaining five tests (absurdities, Block IV examination score, schematic reading, Air Force need achievement and administrative aptitude) did not seem to fall into any meaningful cluster.

The results from all 12 steps of the stepwise regression procedure (Dixon, 1965) are shown in Appendix A of this report. In this section we present the one, four, and eight factor equations so derived. Each of these regression equations is shown in Table 30.

Table 30

Multiple Regression Equations for Steps One, Four, and Eight of Stepwise Regression Procedure (Electronic Principles Course)*

Step	Regression Equation	Multiple R
one	$y' = 65.586 + .899x_5$	.636
four	$y' = 34.843 + .183x_5 + .042x_7 + .287x_{15} + .112x_{16}$	.713
eight	$y' = 40.430 + 1.041x_1 + .066x_3 + .044x_7 - .018x_9 + .176x_{15} + .118x_{16}$	.762

* x_1 = sequential test level
x_3 = schematic reading/component identification
x_5 = conventional test score
x_7 = confidence score (d')
x_9 = civilian need achievement
x_{15} = electronics aptitude
x_{16} = Block IV exam score

Table 30 further supports the prior contention favoring the validity of the novel measures. Note that the one variable equation yielded an r of .64. The Block IV examination did not become included in the equation until step 4.

Variable 5, conventional test score, was eliminated as a predictor in the latter multiple regression equations. This variable was replaced by variable 1, sequential test level. The reader will recall that variable 1 clustered with variable 5 in the cluster analysis. However, the intercorrelation matrix in Table 29 shows that variable 1 is less related to the other variables in the matrix than is variable 5. The variance that variable 1 has in common with the criterion is almost totally unique, while the variance that variable 5 has in common with the criterion can be accounted for by other variables including variable 1.

Variance analytic results for the statistical significance of the linear prediction for steps one, four, and eight are shown in Table 31. All the F tests are statistically significant.

Table 31

Analysis of Variance Summary for the Significance of Linear Prediction for Steps One, Four, and Eight of the Stepwise Regression Procedure

Step		df	Analysis of Variance Sum of Squares	 Mean Square	 F
1	Regression	1	874.288	874.288	26.516*
	Residual	39	1285.907	32.972	
4	Regression	4	1098.570	274.642	9.313*
	Residual	36	1061.625	29.490	
8	Regression	6	1254.142	209.024	7.844*
	Residual	34	906.053	26.649	

*p < .01

Signal Detection

In the present investigation, the signal detection method of scoring the confidence test did not appear to yield a validity coefficient that is appreciably different from the confidence test score, or the conventionally scored test.

As will be recalled, the theory of signal detection provides a method for controlling and measuring the criterion an observer uses in making decisions about signal existence and it provides a measure of the observer's detection sensitivity (d') that is independent of his decision criterion. The sensitive observer is one who differentiates with few errors between signal and noise.

A receiver operator characteristic (ROC) curve depicts d' as a function of hit rate (HR) and false alarm rate (FAR). In the present context, HR is defined as the proportion of time the subject correctly identifies signal when presented with signal plus noise at a confidence level of .68 or higher. (The results of a significance test determined that in a four choice multiple choice item confidence values up to .68 are not significantly different from chance (.25) confidence estimates. Confidence estimates above .68 are considered significant beyond the chance level.) FAR is defined as the proportion of time the subject incorrectly identifies noise as signal when presented with noise only at a confidence level of .68 or higher. When drawing a ROC curve, the HR is depicted along the ordinate, while the FAR is depicted along the abcissa. The ROC curve shown in Figure 25 presents the average d' for subjects scoring the top, middle, and bottom thirds of the criterion group. The reader will note the large degree of differentiation between the three ROC curves indicating an appreciable degree of discrimination. It is possible that test score improvement could be attained by students who score in bottom third if they adjusted their acceptance criterion upward.

Reliability Analysis

The Kuder-Richardson internal consistency reliability formula (formula 20) was applied in all appropriate cases to the electronic principles tests. In some instances (e.g., confidence test scores, cognition of figural systems test) split half methods were utilized because the data could not be adopted to the Kuder-Richardson algorithm. Reliability coefficients were approximated for tests of 25 items (except for the confidence test which already contained 30 items) using the Spearman-Brown formulation. A test length of 25 items was used, inasmuch as this length seemed to be acceptable in view of testing constraints in the military school situation.

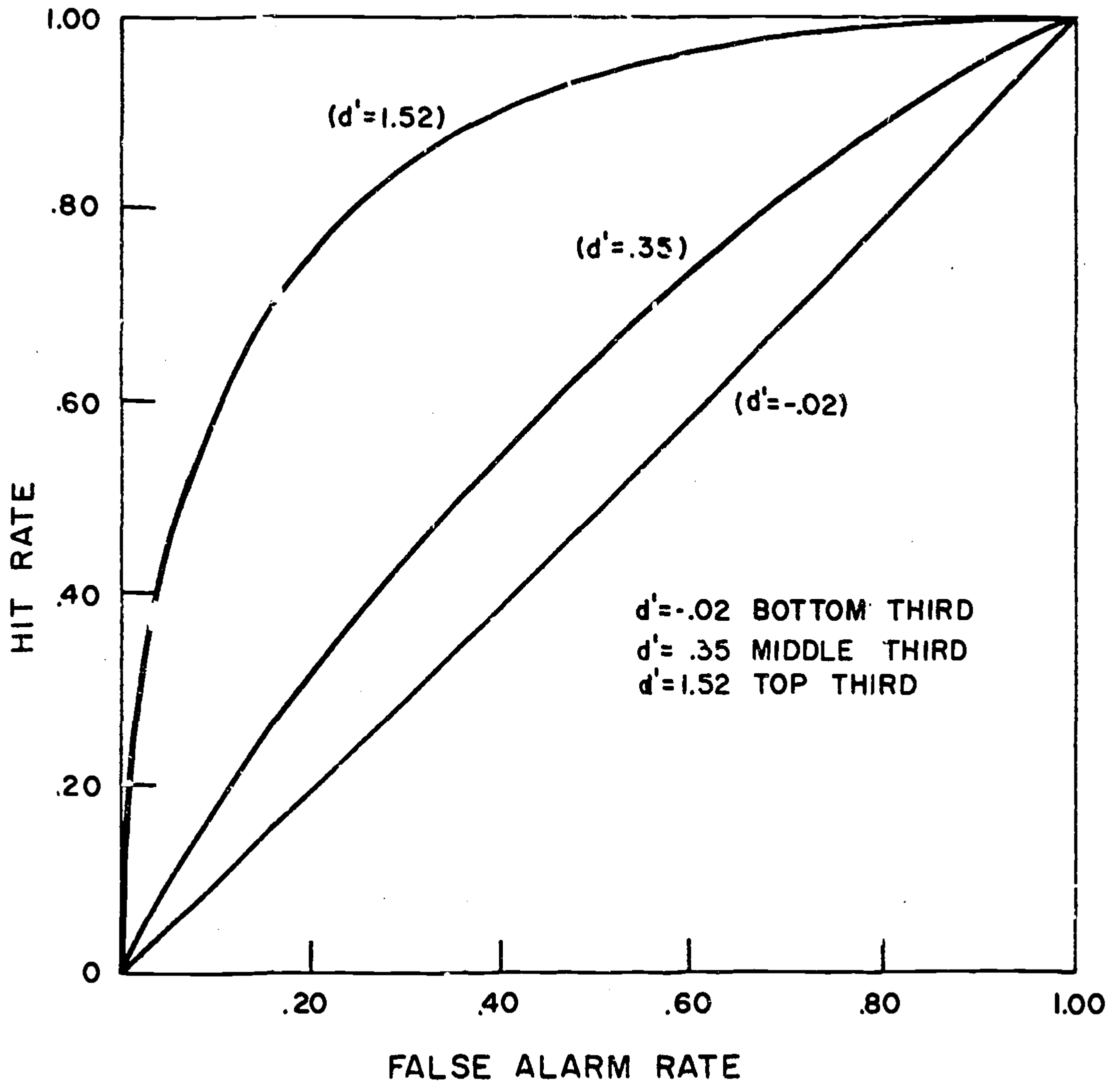


Figure 25. Det c'ion sensitivity (d') for students scoring in the bottom, middle, and top thirds of the criterion.

Table 32 presents the corrected reliability coefficients for the tests taken by the electronic principles students. When comparing tests developed on the basis of items constructed by Applied Psychological Services with tests using electronics school items, it seems as though the reliabilities of the latter are only marginal, while the reliabilities of the former range from marginal to high. These results suggest a need for school examination of the internal consistency reliability of their end of block tests. Reliability coefficients of .6 -.7 are certainly marginal.

Table 32

Reliability Coefficients for Naval Electronics Principles Measures

Test	n	Coefficient
Sequential Test I	3	-
Sequential Test II	13	.82
Sequential Test III	44	.65
Sequential Test IV	8	.68
Sequential Test V	6	.79
Confidence Test Score	44	.60
Conventional Test Score	44	.64
Pictorial Absurdities	44	.85
Pictorial Schematic Reading	42	.67
Cognition of Figural Systems	44	.84
Analogies	44	.88
Psychophysiological Arousal*	30	.99

* Interscorer reliability

Questionnaire Analysis

Following completion of the testing, each student who had participated in the electronics testing completed a structured questionnaire. The questionnaire (presented in Appendix E) sought to determine student reaction to each of the advanced measurement techniques. The 16 items included in the questionnaire concerned the relevance, fairness, preference for, complexity, and interest of the various advanced measurement techniques when compared with the usual multiple choice school tests. The student was requested to indicate his agreement or disagreement with each item on a six point scale. Mean responses, across all students, were then obtained. Table 33 presents the mean value, topic, and question number for the questionnaire data. In developing Table 33, all scales that were presented to the students in the reverse direction (approximately half of the scales were originally reversed in order to avoid the effects of response set and acquiescence) were again reversed to yield consistency within the table and ease of comparison. In addition, the items are regrouped according to test or test type for ease of comparison. Mean values between 3.00 and 4.00 on the scale indicate neutrality of opinion regarding the advanced measurement techniques. Values greater than 4.00 indicate that the students responded positively to the advanced measurement technique as compared with the usual school tests, while values less than 3.00 indicate that the students responded unfavorably to the advanced measurement techniques. Questions concerned with the sequential tests and other conventionally scored tests were not included since they represent the usual multiple choice school tests.

Generally, Table 33 suggests that the more novel approaches (e.g., cognition of figural systems, absurdities) were less preferred than the usual school tests. Conversely, novel approaches which represented refinements of traditional testing procedures (e.g., confidence testing, schematic reading) were considered fairer and more relevant. There are several reasons why the students may have tended not to prefer the more novel testing techniques. First, many individuals will react with fear and dislike when presented with anything new or unusual. They will often feel incompetent and unable to cope with the new situation. Given time and experience, though, attitudes often change.

Table 33

Mean Response to Questionnaire Items for Various Tests

Question Number	Test	Topic	Mean
1	Confidence Test	Fairness	4.49
6	Confidence Test	Preference	3.68
5	Pictorial Absurdities	Preference	2.15
3	Pictorial Schematic Reading	Relevance	4.07
8	Pictorial Schematic Reading	Preference	3.49
10	Cognition of Figural Systems	Preference	1.85
13	Cognition of Figural Systems	Preference	2.51
2	Analogies	Relevance	2.68
7	Analogies	Preference	2.66
4	All Advanced Measurement Techniques	Interest	3.73
9	All Advanced Measurement Techniques	Difficulty	3.68
11	Advanced Measurement Test Directions	Difficulty	5.17
12	All Advanced Measurement Techniques	Interference	2.98
14	All Advanced Measurement Techniques	Interest	3.54
15	All Advanced Measurement Techniques	Complexity	4.02
16	All Advanced Measurement Techniques	Relevance	3.98

Second, the students indicated the experimental testing program to interfer with their daily school schedule, and for this reason, they may have reacted negatively toward some of the more unique measurement techniques. Finally, the novel measures were more abstract and difficult, and they provoked the individual to use higher order thinking processes and concepts. The novel methods did not rely on simple stimulus-response situations as is the case with most multiple choice tests. A test that requires more work on the part of a student is less likely to be preferred.

Interview Analysis

Ten instructors from the electronics school were given individual briefings on the development, rationale, and use of each advanced measurement technique. Each instructor was then systematically interviewed regarding his assessment of the effectiveness of the new technique, its utility in regard to assessing student attainment of training objections, administrative and logistic considerations, and the possible effects of the novel approaches on the role of the instructor and training evaluation specialist. Tables 34-39 present the instructors' evaluation of the effectiveness of each of the new advanced measurement techniques when compared with the usual multiple choice school tests. Effectiveness was rated on a 0-100 scale with a value of 50 representing effectiveness equal to that of multiple choice tests. A value greater than 50 indicated that the instructor considered the advanced measurement technique to be more effective than the usual multiple choice electronics school test, while a value of less than 50 indicated that the instructor thought the advanced measurement technique was less effective than the multiple choice school test. Value judgments between 45 and 55 indicate a neutral judgment.

The results presented in Tables 34-39 indicate that the instructors, as a group, had a favorable opinion of the sequential test (because of its ability to improve the channeling of students), the confidence test (because of the increased information it provides), the figural systems test (because it tests understanding), and the schematic reading test (because a similar format is currently used). The instructors seem to have had a neutral opinion of the analogies test and an unfavorable opinion (for a variety of reasons) of the absurdities test. These two tests also have been shown earlier to exhibit low validity coefficients.

Table 34

Evaluation of Effectiveness of the Sequential Tests as Compared with Multiple Choice Tests by Ten Electronics Instructors at Keesler Air Force Base

Instructor Number	Effectiveness	Comments
1	90	A hell of a lot more effective.
2	75	No comment
3	60	No comment
4	83	This will be a very effective method for channeling students on the basis of test results when we go over to self pacing next year.
5	75	For self pacing it will be better than what we have now.
6	50	Can't see any more or less advantage.
7	75	You can advance him into something else. You're not stuck with one test.
8	50	Same in effectiveness.
9	80	You can find out which students learn the material most quickly and you can get them onto something else. You identify the person who needs help. Now, they teach at the lowest level demotivating many students.
10	40	Must stick to lock step, this is most adaptable to self pacing. Because what we've got now where students are given certain material on a certain day.
Mean	67.80	

Table 35

Evaluation of Effectiveness of the Confidence Test Compared with Multiple Choice Tests by Ten Electronics Instructors at Keesler Air Force Base

Instructor Number	Effectiveness	Comments
1	75	Ease of grading. It does give better picture of how much he knows and removes guess factor. More difficult to grade.
2	70	No comment
3	90	No comment
4	75	If students had time to learn how to use it.
5	75	Might help but don't know if it can completely test a man's knowledge. Would not want to use it instead of straight objective tests. It's better for diagnostic purposes.
6	38	I want to know if the student is sure of the right answer. If he's only partly sure that's not good enough. He's got to be able to perform on the job.
7	75	It might be good with some of the test items, but not with all. Sometimes he must know it all to satisfy a criterion objective.
8	90	It's more important to be sure of an answer than to answer the question right.
9	30	Less effective. Confusion among students. A good student will underrate himself and a poor student will overrate himself.
10	95	Gives you differentiation between students who know the material and those guessing. Gives credit where credit is due.
Mean	71.30	

Table 36

Evaluation of Effectiveness of the Absurdities Test as Compared with Multiple Choice Tests by Ten Electronics Instructors at Keesler Air Force Base

Instructor Number	Effectiveness	Comments
1	35	His job is not to build a circuit, but to work with what he has. He'll work on a complete circuit and troubleshoot on that circuit. Not realistic.
2	25	Too much memorization and recognizing circuitry.
3	65	No comment
4	25	We don't instruct at that high a level. We teach basic principles of electronics. Good in field.
5	50	Similar to troubleshooting in the lab.
6	25	Alone, against standard tests it would be less effective. Students called it "dopey" test.
7	25	They would have to recall what circuit should look like.
8	25	We don't like to show students something that is wrong, because he might retain this. This could be an informative test tool for advanced training not down on the elementary level.
9	40	Too time consuming. Con't use other knowledge. It's a total recall test.
10	35	Good in some situations. Measures observation rather than information.
Mean	40.00	

Table 37

Evaluation of Effectiveness of the Schematic Reading/Component Identification Test as Compared with Multiple Choice Tests by Ten Electronics Instructors at Keesler Air Force Base

Instructor Number	Effectiveness	Comments
1	75	Test followed objectives well. Not too different from the tests we have now. Questions got right to the point.
2	50	About the same as what we are doing now.
3	50	Similar to what is used now.
4	50	It's almost the usual school tests.
5	50	Same
6	55	It's good. Used in part of a system test. Would not want to use it alone. It's more effective in some places. It gets at concepts and helps to test degree of knowledge about the circuit.
7	50	We do about the same thing today, especially when teaching the circuit.
8	50	No comment
9	70	You're going to have to know what component does and current flow. You have to know circuit to be able to connect it and for it to work and this is what they do on the job anyway. Job relevant.
10	70	Because they have to know it. Troubleshooting.
Mean	57.00	

Table 38

Evaluation of Effectiveness of the Cognition of Figural Systems Test as Compared with Multiple Choice Tests by Ten Electronics Instructors at Keesler Air Force Base

Instructor Number	Effectiveness	Comments
1	30	Objective is circuit identification, not building a circuit.
2	25	It's a difficult test considering we have a basic electronics course. We can't teach him that much material in the length of time allotted.
3	70	Good system
4	75	No comment
5	25	No comment
6	75	I like it. A good way to test students in circuits. It tests man's ability to analyze, logical thinking and knowledge at the same time. It should not be used alone, but as part of a test system. A good test.
7	80	They will really have to study a circuit to see how it is wired together. It will require the students to learn more because in the past the teacher told them the relationship between parts. Now they will have to do it.
8	10	We don't teach or have them memorize it. If we emphasized practicing a procedure when they can know the exact arrangement of things, maybe.
9	75	You're going to have to know what component does and current flow. You have to know circuit to be able to connect it and for it to work and this is what they do on the job anyway. Job relevant.
10	55	Identifying schematic is important. This makes sure they know it. It overemphasizes something we don't intend to emphasize.
Mean	57.00	

Table 39

Evaluation of Effectiveness of the Analogies Test as Compared with Multiple Choice Tests by Ten Electronics Instructors at Keesler Air Force Base

Instructor Number	Effectiveness	Comments
1	65	Makes the student stop and think. Can't guess. Will make student sit down and study or he hasn't a chance.
2	25	Won't test what he needs to learn. It's just comparing pictures. Memorization rather than knowing and understanding a circuit.
3	10	Because wouldn't be able to work with it. They don't know that much about it. Too much partial visualization required.
4	70	No comment
5	40	A little harder.
6	25	Use of this test would be limited. May be good for some things like circuit recognition. As one of a gang of tests. O.K. as far as it goes would be less effective than straight multiple choice.
7	50	Correlation between two different things is used today in some way.
8	50	We do this on a very limited scale, presently, but we don't have a test full of this testing approach.
9	50	Too time consuming.
10	85	Some questions were ambiguous. You're not testing reading or wording, only what he knows about electronics.
Mean	47.00	

Table 40 presents the instructors' estimations of the probability that administrative or logistics problems would arise in the use of the advanced measurement techniques. The instructors were asked to estimate the chances in 100 that these kinds of problems would arise; therefore, the higher the mean value shown in the table, the greater the instructors' estimation that that administrative and logistics problems would arise.

Table 40

Estimation by 10 Electronics Instructors of the Probability that Administrative or Logistics Problems Would Arise in the Use of the Advanced Measurement Techniques

Technique	Probability
Sequential Tests	.46
Confidence Tests	.40
Absurdities Test	.43
Schematic Reading/Component Identification Test	.29
Configuration of Figural Systems Test	.62
Analogies Test	.36

Table 40 reveals that the schematic reading test estimate is the only one with a low mean probability value. All of the remaining tests exhibit roughly twice the problem probability of the schematic reading test. The schematic reading test is similar in format to an already employed test. The particular logistical and administrative problems foreseen by the instructors centered on three areas: (a) need for more personnel, (b) grading and scoring problems, and (c) time consumption. It seems, however, that these problems must be weighted against any increased value for the novel measures, if employed.

A final question asked the instructors about the effects of the advanced measurement techniques on themselves and the training evaluation specialist. The most pervasive effect of the advanced measurement techniques, as indicated by the instructors, was that they would necessitate a change in teaching content in the direction of more circuit analysis. This was especially true of the more abstract and difficult tests (e.g., figural systems, analogies, and absurdities).

Cost/ Effectiveness Analysis

The technique employed for assessing cost/ effectiveness involved application of the following formula:

$$\text{Cost/ Effectiveness} = \left[\frac{DC + AC + SC}{N}\right] \times (1-r)^2 \text{ or } \frac{1}{S}$$

where: DC = Direct costs
AC = Administrative costs
SC = Support costs
N = Number of students evaluated
r^2 = Variance common to test and criterion
S = Number of course improvement suggestions elicited

This formula yields an inverse index, i.e., a high cost/ effectiveness index designates less cost/ effectiveness.

Direct costs can be considered to include the total dollar cost of the necessary man hours for test construction plus the cost of materials. Administrative costs include facility and other indirect costs. These are proportional to direct costs and for the present purposes can be considered to equal direct costs. Support costs include such factors as support equipment, replacement, test administration time, and test scoring time. As direct costs, administrative costs and support costs increase the cost/effectiveness of the test decreases. The variance common to the test and the criterion indicates the degree of relationship between the test and the criterion. The number of course improvement suggestions elicited is a function of the test's intercorrelation with other predictors, as well as the test's validity coefficient. The greater the amount of unique variance in the criterion that the test accounts for, the greater the amount of training feedback information elicited by that test.

In the present analysis, the test's validity coefficient was employed rather than the number of course improvement suggestions elicited. As the validity coefficient increases the cost/ effectiveness value, as calculated, decreases. Table 40 presents "best estimates" of the yearly costs for each factor in the cost/ effectiveness equation for each test, including the usual Block IV test. We have assumed that N is a constant value of 300 students per year. The auxiliary measures, such as achievement motivation and physiological arousal have not been considered, inasmuch as their costs are minimal and because many of the variables in the cost/benefit equation do not apply to them. Additional assumptions are:

1. each sequential test item contains 12 items
2. 300 students will complete 12 sequential test items
3. 150 students will complete 24 sequential test items
4. 75 students will complete 36 sequential test items
5. the test length of the confidence test and its variations is 30 test items
6. the figural systems, absurdities, analogies, and schematic reading tests contain 25 items
7. four forms of each test will be constructed
8. floor space costs $5 a square foot per year and is the same for each test
9. maintenance is $800 per year
10. duplication costs $0.01 per page
11. typing or test preparation costs $4 per page
12. scoring costs and test administration costs are $6 per hour

The results of the cost/effectiveness analysis are presented in Table 41, which suggests that the confidence scoring technique and the signal detection modification of the confidence scoring technique were the most cost effective of the advanced measurement techniques. The schematic reading, analogies, figural systems, sequential, and Block IV tests appear to be moderately cost effective. The absurdities test was least cost effective, probably because of its low correlation with the criterion.

Table 41

Cost/ Effectiveness Factors for Each of the Advanced Measurement Techniques and for the Block IV Electronics Test for One Year

Test	Direct Costs	Administrative Costs	Support Costs	Total Costs	Number of Students	Common Variance	Cost/Effectiveness Index
Sequential	3038	3038	6412	12488	300	.18	34.29
Confidence	1254	1254	6359	8667	300	.33	19.96
Signal detection (d')	1254	1254	6368	8676	300	.41	17.44
Absurdities	2481	2481	6460	11422	300	.00	38.04
Schematic reading	2481	2481	6334	11296	300	.08	34.49
Figural systems	2511	2511	6858	11880	300	.14	33.90
Analogies	2496	2496	6389	11381	300	.08	34.94
Block IV examination	2493	2493	6547	11553	300	.24	29.25

Smaller values indicate greater cost/ effectivness.

Administrative Specialist Measures

Administrative Specialist Correlational Analysis

As with the electronic principles course, each of the administrative specialist advanced measurement techniques was considered as a predictor of school performance in a concurrent validity study. Also included as predictors in this portion of the analysis were the auxiliary measures and the scores on the Block II test in the administrative Specialist School. Three criteria of school performance were used:

1. average of course block grade with the effects of the Block II grades removed

2. total self-paced hours to compute administrative specialist course excluding the hours taken to complete Blocks I and II

3. average of course block grade with the effects of the Block II grades removed and normalized by the square root of the total self-paced hours to complete the administrative course excluding the hours taken to complete Blocks I and II

The last criterion was included to allow a measure of performance which includes the amount learned tempered by the learning time. The square root function was included to reflect the negative accelerated characteristics usually associated with learning curves. (If the square root transformation were not performed most of the variance of this criterion measure would have been attributable to learning time which was also used as a criterion measure.)

The stepwise regression (Dixon, 1965) method of analysis was also applied to the administrative specialist data. Prior to the completion of the stepwise regression analysis, in order to eliminate decimals and negative numbers, several of the advanced measurement technique scores were either multiplied by a constant or had a constant added. Those measurement techniques so involved, along with their respective additive and/or multiplicative constants, are shown below in Table 42.

Table 42

Additive and/or Multiplicative Constants for Certain Administrative Specialist Advanced Measurement Techniques

Technique	Additive Constant	Multiplicative Constant
Partial Knowledge	4.00	
Thurstone Typing Samples (all)	250.00	
Absurdities	7.00	
Signal Detection	12.00	10.00
Achievement Motivation	32.00	

The means and standard deviations of the advanced measurement techniques, the auxiliary measures, the Air Force aptitude variables, the Block II score, and the criterion variables for the 31 students sampled are presented in Table 43. The intercorrelation matrix for these same variables and subjects is presented in Table 44.

Table 43

Means and Standard Deviations of the Advanced Measurement Techniques, the Auxiliary Measures, the Air Force Aptitude Variables, the Block II Score, and the Three Criterion Variables for 31 Students Enrolled in the Administrative Specialist Course

Variable	Mean	Standard Deviation
Partial Knowledge	31.032	11.563
Signal Detection (d')	115.484	46.586
Technical Words	18.677	4.700
Typing I	264.710	17.162
Typing II	216.645	58.756
Typing III	243.387	26.267
Typing IV	241.935	25.857
Absurdities	12.258	4.604
Need Achievement (Civilian)	132.484	52.789
Need Achievement (Air Force)	129.484	47.109
Psychophysiological Arousal	62.226	28.036
General Aptitude	51.452	19.925
Mechanical Aptitude	42.097	20.280
Administrative Aptitude	57.581	17.023
Electronics Aptitude	48.065	19.394
Block II Examination Score	83.355	11.960
Final Course Average	79.903	10.537
Course Hours	119.968	54.682
Course Average/ $\sqrt{\text{Course Hours}}$	7.983	2.037

Table 44

Intercorrelation Matrix* of the Advanced Measurement Techniques, the Auxiliary Measures, the Air Force Aptitude Variables, the Block II Score, and the Three Criterion Variables for 31 Students Enrolled in the Administrative Specialist Course

Variables	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Partial Knowledge	972	416	343	322	270	-080	136	304	042	394	354	474	301	408	486	638	-367	464
2. Signal Detection		409	312	250	249	-098	-098	197	-063	380	330	400	301	375	451	627	-337	440
3. Technical Words			268	398	341	214	363	000	-103	211	381	430	344	435	392	645	-677	735
4. Typing I				432	202	032	-028	281	004	010	187	397	058	156	171	157	-362	371
5. Typing II					657	393	273	-097	-142	-122	173	205	105	201	265	104	-339	226
6. Typing III						405	146	-100	-143	190	157	261	207	434	182	260	-254	251
7. Typing IV							117	-535	-290	-213	-142	-100	000	-104	062	041	-191	108
8. Pictorial Absurdities								-054	-180	040	397	460	414	295	412	418	-275	403
9. Need Achievement (Civilian)									711	353	098	502	003	259	045	256	-084	208
10. Need Achievement (Air Force)										123	023	340	-179	124	-047	102	164	014
11. Psychophysiological Arousal											124	212	-008	200	076	442	-207	346
12. General Aptitude												704	765	663	471	441	-162	333
13. Mechanical Aptitude													650	727	533	656	-238	525
14. Administrative Aptitude														808	526	452	-211	348
15. Electronics Aptitude															472	516	-292	438
16. Block II Exam. Score																613	-392	519
17. Final Course Average																	-437	-084
16. Course Hours																		740
19. Course Average/ $\sqrt{\text{Course Hours}}$																		

* Decimal points have been eliminated.

The obtained validity coefficients for final course average and final average divided by the square root of hours to complete course are generally quite acceptable. Raw hours to complete course were not indicated to be a criterion which can be predicted by the advanced measurement techniques, the Air Force aptitude tests, the usual school Block II examination, or the auxiliary measures. In terms of the two predictable criteria, the technical words, absurdity, partial knowledge, and d' approaches demonstrated validity coefficients which ranged from .42 to .74. These concurrent validity coefficients are exceptionally high for this type of situation. The Thurstone scaled typing tests individually showed little validity as did the measures of need achievement. Some promise was indicated for the psychophysiological arousal measure in this context. To collapse systematically the intercorrelational matrix on the basis of the intercorrelations among the predictor variables, a cluster analysis was performed. The results of the cluster analysis are shown in Table 45.

The results of the cluster analysis demonstrate that the largest cluster consists of the four Air Force aptitude tests and the Block II exam score. The second cluster contains the two partial knowledge tests and the technical words test. The two need achievement indices comprise the third cluster. The second, third, and fourth typing tests constitute the fourth cluster. The first typing test, psychophysiological arousal, and the absurdities tests remained independent of the four major clusters.

Table 44 indicates a positive relationship ($r = .35$) between achievement motivation (civilian) and the psychophysiological arousal index. The electronics data, though, indicate these variables to be inversely related ($r = -.44$). It is noted that the mean general aptitude test score for the electronics sample (78.8) was significantly higher than that of the administrative sample (51.5). This suggests that aptitude may moderate the relationship between arousal and achievement motivation. That is, physiological arousal and achievement seem inversely related in high aptitude groups and positively related in low aptitude groups.

TABLE 45

CLUSTER ANALYSIS OF 16 PREDICTOR VARIABLES DERIVED FROM THE SCORES OF 31 ADMINISTRATIVE SPECIALIST STUDENTS

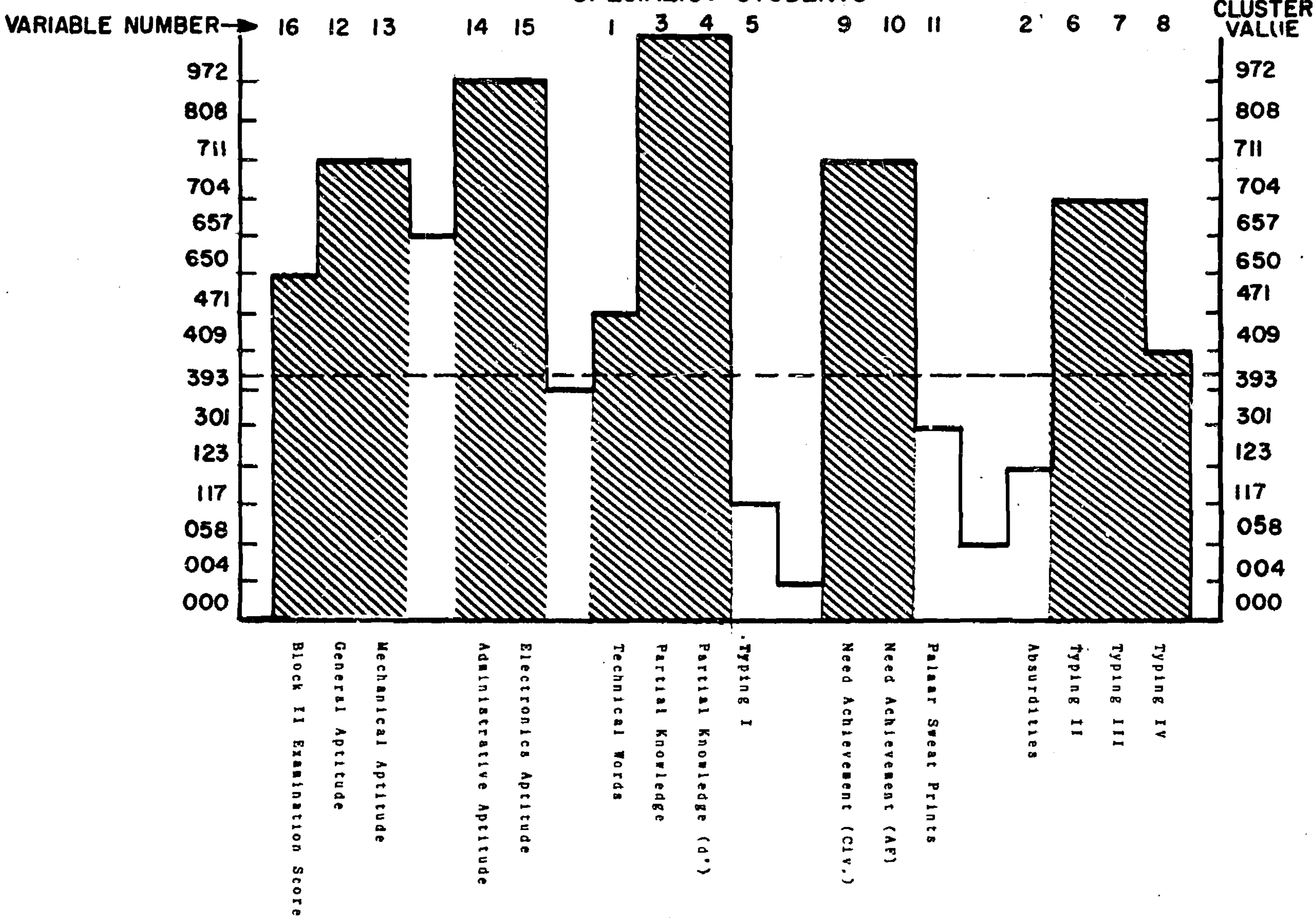


Table 46 presents the means and t values for the four aptitude tests across the electronics and administrative groups.

Table 46

Means and t-Ratios for the Four Air Force Aptitude Tests Across the Administrative and Electronics Groups

Test	Administrative Mean	Electronics Mean	t-Ratio
General	51.45	78.78	6.65*
Mechanical	42.10	77.07	8.26*
Administrative	57.58	71.10	3.22*
Electronic	48.07	86.34	10.51*

*$p < .01$

Table 46 indicates a substantial difference in aptitude across the administrative and electronics samples.

The first 12 steps of the stepwise regression procedure for each of the criteria (Dixon, 1965) are shown in Appendices B, C, and D of this report. Steps 1, 4, and 8 are presented for each of the criteria in Tables 47-49.

Table 47

Multiple Regression Equations for Steps One, Four, and Eight of Stepwise Regression Procedure for the Prediction of Final Course Average (Administrative Specialist Course)*

Step	Regression Equation	Multiple R
one	$y' = 65.552 + .341x_{13}$	.656
four	$y' = 53.415 + .961x_1 + .081x_4 - .042x_6 + .195x_{13}$	.851
eight	$y' = 25.870 + .751x_1 + .312x_2 + .070x_4 - .051x_6 + .061x_8 + .067x_{11} + .133x_{13} + .182x_{16}$	.899

* x_1 = Technical words

x_2 = Absurdities

x_4 = Partial knowledge (d')

x_6 = Typing II

x_8 = Typing IV

x_{11} = Palmar sweat prints

x_{13} = Mechanical Aptitude

x_{16} = Block II examination score

Table 48

Multiple Regression Equations for Steps One, Four, and Eight of Stepwise Regression Procedure for the Prediction of Course Hours (Administrative Specialist Course)*

Step	Regression Equation	Multiple R
one	$y' = 266.967 - 7.870x_1$	.677
four	$y' = 535.339 - 7.276x_1 - .819x_5 + .707x_{13} - 1.108x_{16}$	.741
eight	$y' = 742.568 - 6.712x_1 - 2.832x_2 - .986x_5 - .389x_8 - .426x_9 + 2.065x_{13} - .747x_{15} - 1.225x_{16}$	.798

* x_1 = Technical words

x_2 = Absurdities

x_5 = Typing I

x_8 = Typing IV

x_9 = Civilian need achievement

x_{13} = Mechanical Aptitude

x_{15} = Electronics Aptitude

x_{16} = Block II examination score

Table 49

Multiple Regression Equations for Steps One, Four, and Eight of Stepwise Regression Procedure for the Prediction of Final Course Average Divided by the Square Root of Course Hours (Administrative Specialist Course)

Step	Regression Equation	Multiple R
one	$y' = 20.342 + 3.185x_1$	.735
four	$y' = -66.890 + 2.342x_1 + .214x_5 + .153x_{11} + .444x_{16}$	.820
eight	$y' = -95.736 + 2.524x_1 + .910\ x_2 + .339\ x_5 = .069\ x_6 + .033x_{10} + .127x_{11} - .141x_{12} + .451x_{16}$	.852

* x_1 = Technical words

x_2 = Absurdities

x_5 = Typing I

x_6 = Typing II

x_{10} = Air Force need achievement

x_{11} = Palmar sweat prints

x_{12} = General aptitude

x_{16} = Block II examination score

Selection of the most useful prediction scheme from among those presented in Tables 47-49 rests with the goals of the user. If final course average prediction represents the goal of the user, then the equations shown in Table 47 would be used. If, on the other hand, time to complete the course (with a passing grade) is the goal, then the equations presented in Table 48 would be used. If prediction of time to complete the course in conjunction with grade represents the user's goal, then the equations given in Table 49 would be used.

Several of the predictors appear in all three criterion situations. These predictors are technical words, absurdities, and Block II exam score. Technical words and absurdities were two of the advanced measurement techniques. Six other predictors appeared in two of the three criterion situations. These were:(a) typing I, (b) typing II, (c) typing IV, (d) psychophysiological arousal, (e) mechanical aptitude, and (f) Block II examination score.

We note the absence of the administrative aptitude test in any of the regression equations.

Summaries of analyses of variance for the significance of linear prediction for steps 1, 4, and 8 of each prediction scheme are shown in Tables 50-52. Tables 50-52 indicate statistically significant F tests in all instances.

Table 50

Analysis of Variance Summary for the Significance of Linear Prediction for Steps One, Four, and Eight of the Stepwise Regression Procedure Predicting Final Course Average

			Analysis of Variance		
Step		df	Sum of Squares	Mean Square	F
1	Regression	1	1433.933	1433.933	21.924*
	Residual	29	1896.777	65.406	
4	Regression	4	2411.525	602.881	17.053*
	Residual	26	919.185	35.353	
8	Regression	8	2693.012	336.626	11.613*
	Residual	22	637.698	28.986	

*p < .01.

Table 51

Analysis of Variance Summary for the Significance of Linear Prediction for Steps One, Four, and Eight of the Stepwise Regression Procedure Predicting Course Hours

			Analysis of Variance		
Step		df	Sum of Squares	Mean Square	F
1	Regression	1	41054.737	41054.737	24.473*
	Residual	29	48648.230	1677.525	
4	Regression	4	49210.313	12302.578	7.899*
	Residual	26	40492.655	1557.410	
8	Regression	8	57151.536	7143.942	4.828*
	Residual	22	32551.432	1479.611	

*p < .01.

Table 52

Analysis of Variance Summary for the Significance of Linear Prediction for Steps One, Four, and Eight of the Stepwise Regression Procedure Predicting Final Course Average Divided by the Square Root of Course Hours

Step		df	Analysis of Variance Sum of Squares	Mean Square	F
1	Regression	1	67238.834	67238.834	34.093*
	Residual	29	57194.841	1972.236	
4	Regression	4	83598.072	20899.518	13.302*
	Residual	26	40835.604	1570.600	
8	Regression	8	90257.033	11282.129	7.262*
	Residual	22	34176.643	1553.484	

* $p < .01$.

Signal Detection

In the administrative specialist phase of the correlational analysis, as with the electronic principles analysis, the signal detection method of scoring the partial knowledge test did not seem to yield a validity coefficient that is appreciably different from the partial knowledge test score.

A receiver operator characteristic (ROC) curve depicts d' as a function of hit rate (HR) and false alarm rate (FAR). In the present context, HR is defined as the proportion of time the subject correctly identifies signal when presented with signal plus noise. FAR is defined as the proportion of time the subject incorrectly identifies noise as signal when presented with noise. The ROC curve shown in Figure 26 presents the average d' for subjects scoring in the top, middle, and bottom thirds on the final course average criterion. There is considerable difference between the ROC curves for the lowest criterion group and the other two groups.

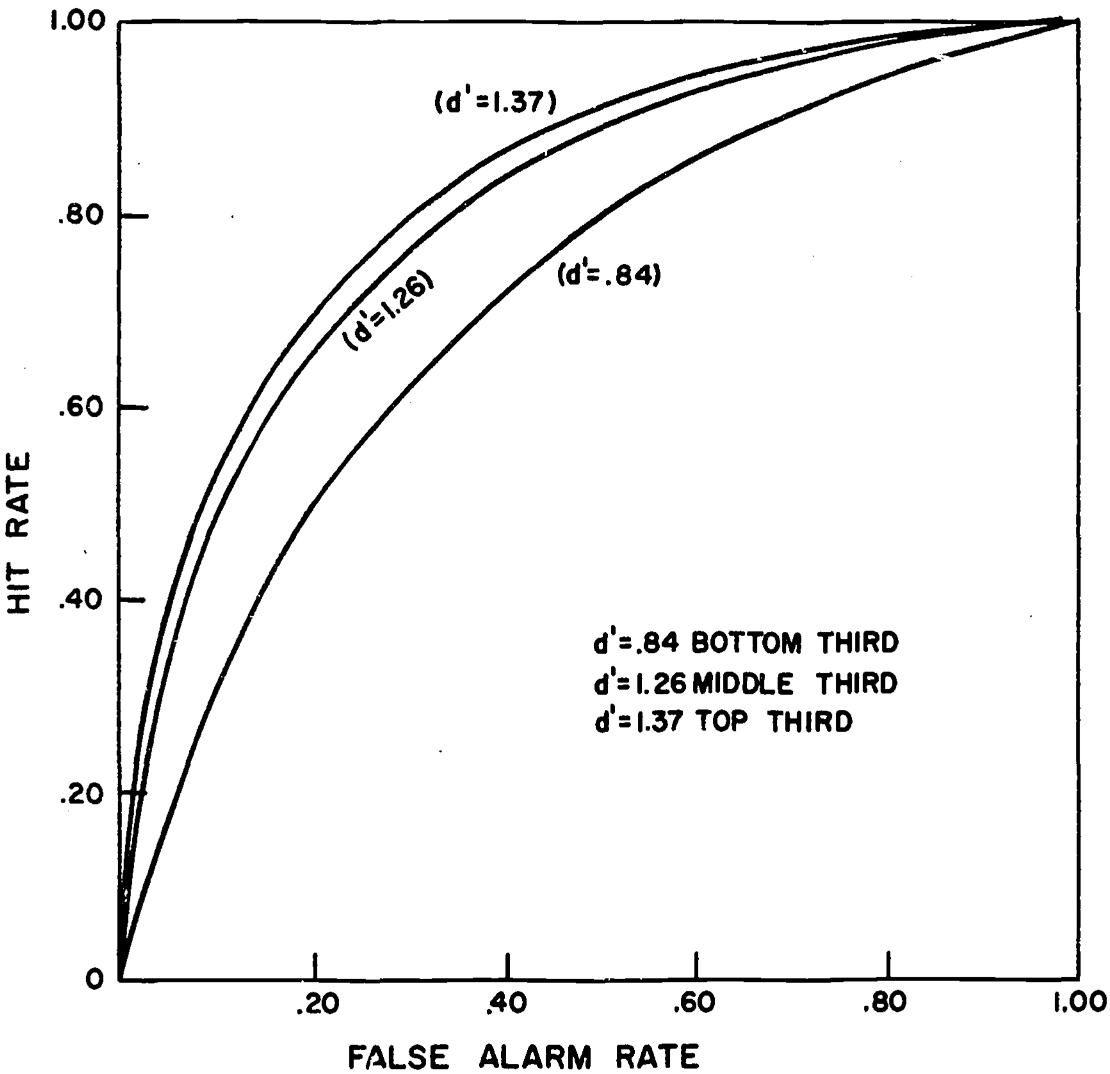


Figure 26. Detection sensitivity (d') for students scoring in the bottom, middle, and top thirds of the criterion.

Reliability Analysis

As was the case with the electronic principles test, the Kuder-Richardson 20 internal consistency reliability formula was applied in all appropriate cases to each of the administrative specialist tests. Only the obtained reliability coefficient for the absurdity test was corrected through the Spearman-Brown formula. The remaining tests, partial knowledge and technical words, met the 25 test item requirement. Table 53 presents the obtained reliability coefficients for the tests taken by the administrative specialist students. Internal consistency reliabilities for the four typing tests could not be calculated given the nature of the typing data. Instead, the number of reversals from the Thurstone scale was considered to be our index of typing inconsistency. The number of reversals was divided by the total number of possible reversals. This proportion was subtracted from one to give an index of reliability.

The data included in Table 53 indicate the partial knowledge and the typing scores to possess unacceptable reliability, while the technical words and absurdity tests yielded reliability indices in the acceptable range.

Table 53

Reliability Coefficients for Three Novel Administrative Specialist Measures

Test	n	Coefficient
Partial knowledge	37	.50
Technical words	38	.77
Absurdities	31	.77
Typing	38	.54

Questionnaire Analysis

After testing was completed, each administrative student completed a structured questionnaire concerned with the desirability or undesirability of each of the advanced measurement techniques. The 12 items included in the questionnaire inquired into the examinees' opinion regarding the relevance, fairness, preferability, complexity, and interest of the advanced measurement techniques as compared with the usual multiple choice school tests. In the questionnaire, the student was asked to indicate his agreement or disagreement with each statement along a six point scale. The response means were then calculated for all students. Table 53 presents the mean value, topic, and question number for the aforementioned questionnaire data. All scales that were presented to the students in the reverse direction were again reversed for ease of comparison and consistency. In addition, the questions have been regrouped according to test or test type. As was the case with the electronic tests, values between 3.00 and 4.00 on the scale indicate neutrality of opinion regarding the new advanced measurement techniques. Values greater than 4.00 indicate that the subjects responded positively to the advanced measurement techniques over the usual school tests, while values less than 3.00 indicate that the subjects responded unfavorably to the advanced measurement techniques.

Several conclusions can be derived from the data presented in Table 54:

1. multiple choice tests were preferred over absurdities tests

2. the advanced measurement technique test directions were easy to understand

3. the advanced measurement technique experimental program was thought to interfere with the students' schedule

Table 54

Mean Response to Questionnaire Items for Various Tests

Question Number	Test	Topic	Mean
1	Partial knowledge	Preference	3.35
2	Absurdities	Preference	1.29
3	Technical words	Relevance	4.48
4	Partial knowledge	Preference	1.26
5	Absurdities	Interest	3.45
6	Technical words	Relevance	3.52
7	All Advanced Measurement Techniques	Interest	3.35
8	All Advanced Measurement Techniques	Difficulty	3.94
9	All Advanced Measurement Test Directions	Difficulty	4.10
10	All Advanced Measurement Techniques	Interference	1.90
11	All Advanced Measurement Techniques	Interest	3.10
12	All Advanced Measurement Techniques	Relevance	3.84

Again, it seems as though the more novel approaches to test development (e.g., absurdities, partial knowledge) were less prefered than the more traditional approaches. Also, there appeared to be a tendency for the students to acquiesce when responding to some of the questions. For instance, questions 1 and 5, which were phrased positively toward the advanced measurement techniques, received neutral responses. Conversely, questions 2 and 4, which were phrased negatively toward the advanced measurement techniques, elicited unfavorable responses. In effect, it seems that the items partially persuaded the students to agree regardless of the attitude toward the advanced measurement techniques. Finally, some of the students may have responded negatively toward the advanced measurement techniques because of the interference of the testing programs with their schedule.

Interview Analysis

Ten instructors from the administrative specialist school were given individual briefings on the development, rationale, and use of each advanced measurement technique. Each instructor was then systematically interviewed regarding his assessment of: (1) the effectiveness of the new technique when compared with the old, (2) the utility of each technique relative to assessing student attainment of training objectives, (3) administrative and logistic problems relative to each technique, and (4) the possible effects of each technique on the role of the instructor and training evaluation specialist. Tables 55-58 present the instructors' evaluation of the effectiveness of each of the new advanced measurement techniques when compared with the usual multiple choice type of test. Effectiveness was rated along a scale which ranged from 0 to 100, with a value of 50 representing effectiveness equal to that of multiple choice tests. A value above 50 indicates that the instructor considered the advanced measurement technique to be more effective than the multiple choice school tests, while a value below 50 indicated that the instructor thought the advanced measurement technique was less effective than the multiple choice school tests. Evaluations of 45 to 55 indicate a neutral judgment effectiveness when comparing the advanced measurement techniques to the multiple choice school tests.

Table 55

Evaluation of Effectiveness of the Partial Knowledge Test as Compared with Multiple Choice Tests by Administrative Specialist Instructors at Keesler Air Force Base

Instructor Number	Effectiveness	Comments
1	50	No comment
2	20	Makes them think too much. Too used to putting down a right answer. Too much of a change
3	75	No comment
4	75	No comment
5	75	No comment
6	75	Builds his morale to know he has something right
7	90	You find out what the student knows. A better view of his understanding of the subject
8	75	Anything that rewards the student for a partial answer motivates the student and will help his job
9	75	You get a chance to tell what you do know. Sort of like an essay test
10	90	The method gives man credit for what he doesn't know. Honesty. Eliminates guesswork. A valid test.
Mean	70.00	

Table 56

Evaluation of Effectiveness of the Technical Words Test as Compared with Multiple Choice Tests by Ten Administrative Specialist Instructors at Keesler Air Force Base

Instructor Number	Effectiveness	Comments
1	50	No comment
2	65	They'll learn more
3	50	No comment
4	75	No comment
5	75	No comment
6	50	About the same
7	50	Same
8	85	A lot don't know terminology
9	50	Same as school tests
10	92	Better, because if a man knows a term he knows something about it. You're pulling out what the man knows
Mean	64.20	

Table 57

Evaluation of Effectiveness of the Thurstone Typing Sample as Compared with the Usual School Tests by Ten Administrative Specialist Instructors at Keesler Air Force Base

Instructor Number	Effectiveness	Comments
1	90	No comment
2	10	You can tell a bad from a good typist without it.
3	75	No comment
4	100	You're able to analyze the student and what his problems are.
5	75	No comment
6	50	About the same.
7	60	A little more effective. We already have a tool for diagnosing typing problems. Diatype can be used for this purpose.
8	95	There's no consistency in diagnosing what each individual thinks of a student. We need to reduce the causes of difficulty.
9	75	You could measure the point to where some-one is efficient. He can do it this far.
10	75	Better than what we've got.
Mean	70.50	

Table 58

Evaluation of Effectiveness of the Absurdities Test as Compared with Multiple Choice Tests by Ten Administrative Specialist Instructors at Keesler Air Force Base

Instructor Number	Effectiveness	Comments
1	60	No comment
2	45	It would have to be basic.
3	65	We've had it before.
4	75	No comment
5	25	Wouldn't give him knowledge he needs to do it.
6	60	No comment
7	25	The student might be able to know what's right and what's wrong, but not be able to pick it out of a picture. He may overlook it. Too much room for error.
8	50	About the same.
9	0	You know material and still not be able to pick out what's wrong.
10	100	With a picture you can see what's going on. It's realistic.
Mean	50.50	

The interview data presented in Tables 55-58 indicate that the instructors had a favorable opinion of the partial knowledge test, the technical words test, and the Thurstone typing samples. The instructors had a neutral opinion of the absurdities test.

Table 59 presents the instructors' estimation of the probability that administrative or logistics problems would arise in the use of the advanced measurement techniques. The instructors were asked to estimate the chances in 100 that administrative and logistics problems would arise. Accordingly, the higher the mean value shown in Table 59, the greater the instructors' estimation that administrative and logistics problems would arise.

Table 59

Estimation by Ten Administrative Instructors that Administrative or Logistics Problems Would Arise in the Use of the Advanced Measurement Techniques

Technique	Probability
Partial knowledge	.21
Technical words	.15
Thurstone typing samples	.15
Absurdities	.23
Traditional multiple choice tests	.14

Table 59 indicates rather conclusively that the administrative instructors considered the new tests to involve very few administrative and logistics problems.

The final interview question concerned the possible effects of the advanced measurement techniques on the role of the instructor and the training evaluation specialist. The most consistent effect, as seen by the instructors, was a need for more time when dealing with the advanced measurement techniques. A few of the instructors said that their own knowledge would have to be increased in order to use the advanced measurement tests.

Cost/Effectiveness Analysis

The same cost/effectiveness formula and definitions as employed for the electronic principles cost/effectiveness analysis were used for the cost/effectiveness analysis of the administrative tests.

Table 60 presents the employed estimate of the yearly costs for each factor in the cost/effectiveness equation for each test including the Block II test. The achievement motivation and physiological arousal measures were not considered, since their costs are negligible, and because many of the variables in the cost/effectiveness equation do not apply to them. In computing these cost/effectiveness indices:

1. the validity coefficients of final course average and course hours were the only ones used since they are essentially uncorrelated while the final average divided by the square root of course hours is correlated with each of the other two criteria

2. the validity coefficients relating the highest typing score obtained with the criteria were used in computing cost/effectiveness of the typing tests.

Additionally, the following assumptions were made:

1. 300 students will complete the administrative course in one year
2. 4 Thurstone typing samples take 10 minutes to score
3. each test, except for the typing tests and the Block II examination, contain 25 items
4. the Block II examination contains 50 items
5. each test, except for the typing tests, has four examinations
6. floor space cost is $5 a square foot per year and is the same for each test
7. use of 50 typewriters, for the typing tests, for one year costs $300
8. paper costs $0.01 per sheet and a pencil costs $0.05
9. maintenance is $800 per year
10. duplication costs $0.01 per page
11. typing or test preparation costs $4 per page
12. scoring costs and test administration costs are $6 per hour

The cost/benefit indices for the various tests are presented in Table 60. In Table 60, low indices indicate a relatively greater cost/benefit.

Table 60 demonstrates that the partial knowledge test, the signal detection modification of the partial knowledge test, and the technical words test were the most consistently cost/effective of the administrative advanced measurement techniques. The Thurstone typing test, the absurdities test, and the Block II examination were only moderately cost effective.

Table 60

Cost/Effectiveness Factors for Each of the Advanced Measurement Techniques and the Block II Test Calculated Separately for the Final Course Average Criterion and the Course Hours Criterion for One Year

Test	Direct Costs	Admin. Costs	Support Costs	Total Costs	Number of Students	Final Course Average Common Variance	Course Hours Common Variance	Cost/Benefit (Course Aver.)	Cost/Benefit (Course Hours)
Partial Knowledge	1227	1227	6159	8613	300	.40	.13	17.23	24.98
Partial Knowledge (d')	1227	1227	6234	8688	300	.39	.11	17.09	25.77
Technical Words	1227	1227	6009	8463	300	.42	.46	17.43	16.23
Typing	430	430	6518	7378	300	.00	.14	24.59	21.15
Absurdities	1297	1297	6563	9157	300	.17	.08	25.33	28.08
Block II Examination	2439	2439	6218	11096	300	.38	.15	22.93	31.44

IV. DISCUSSION, SUMMARY, AND CONCLUSIONS

The present study possessed two interrelated but separate purposes. The first purpose involved exploring methods for coherently integrating the various steps in the Instructional System Development (ISD) cycle of the Air Force. The second purpose involved an investigation of the potential of advanced testing concepts in the Air Force technical training context.

ISD Integration

The ISD cycle involves a structured set of steps which are involved in the development of each Air Force technical training course. These steps are:

1. analysis of system requirements
2. definition of educational or training requirements
3. development of objectives and tests
4. planning, development, and validation of instruction
5. conducting and evaluating instruction

Working with available job analytic data, multidimensional scaling analysis and cluster analysis were demonstrated to provide comparable results in terms of organizing job task statements into coherent entities. It seems that course training requirements and objectives can similarly be derived in terms of these entities (factors or clusters). The present work demonstrated the applicability of these entities to test development. If course content were organized around these entities and post-training evaluations were similarly rooted, the needed integrating core would be provided throughout the entire ISD cycle. Certainly, from the points of view included in the present study, there is little to preclude the use of the multidimensional scaling or the cluster analytic methods for providing the required integrating nexus. Unless such a common core is provided throughout, ambiguities in and misinterpretations of the job analysis can tend to introduce early error that becomes amplified in the system. Moreover, the factors or clusters provide a common definitional substrate for each ISD step and as such can help to minimize mismatches at the various nodal points. Without such a common factor approach, it seems apparent that the persons responsible for each stage of course development through the ISD system will develop different goals and definitions. The result is apt to be a number of different perceptions of the entity. In this case, a disorganized end result is apt to result.

The present work also demonstrated that measurement methods, based on the derived factors, can yield valid, reliable, and cost/effective student achievement scores. Implementation of the use of certain of the advanced measurement techniques might involve certain qualitative administrative problems. However, these disadvantages were not reflected in the cost/benefit analyses which, in general, supported the advanced concepts.

The use of all of the advanced measurement techniques here investigated cannot be defended. However, for the electronics tests, the data suggest the following to warrant consideration: sequential testing, figural systems, confidence testing, and d'. For the administrative specialist training consideration seems warranted of: technical words, absurdities, partial knowledge, and d'.

It was thought that the advanced measurement methods, along with the aptitude variables and the auxiliary measures could be used as predictors of final course average. We have already indicated that the end of course criterion is somewhat contaminated since the block tests and the Air Force Aptitude measures involve the same intellectual and test taking abilities that are found in the criterion measure. There is a considerable proportion of variance common to the Air Force tests and the criterion measure that can be attributed to factors other than aptitude or course knowledge (e.g., reading ability, cognitive style, prior testing experience, etc.). Followup studies, using an on-the-job performance criterion would constitute a more defensible validational paradigm.

We hypothesized earlier that most of the advanced measurement techniques employed were involved with unique cognitive functions not found in the aptitude variables or the block tests. In fact, one of two sets of results seem to support this hypothesis. First, the cluster analysis demonstrated that the Air Force Aptitude tests and the block tests tend to fall into the same cluster. In addition, those advanced measurement techniques using course items (e.g., sequential tests, confidence tests) also seem to be highly correlated. Second, in the regression analysis itself, fewer of the highly correlated Air Force variables entered into the regression equation than would be expected considering their correlations with the criterion.

Of the six electronic principles predictors in the eighth regression step (it will be recalled that the computer program eliminated one variable from the regression equation at the eighth step), four were advanced measurement techniques--sequential test level, schematic reading/component identification, signal detection score on the confidence test, and civilian need achievement. The two Air Force predictors in this step were the electronics aptitude and the Block IV test score. These results also indicate the utility of the advanced measurement techniques for predicting final average in the electronic principles course.

When predicting final course average for the administrative specialist course, six of the predictors in the eight variable equation were part of the advanced measurement battery. The six predictors included were the technical words test, the absurdities test, the signal detection method of scoring the partial knowledge test, two of the Thurstone typing samples, and the measure of psychophysiological arousal. The two Air Force predictors at the eighth regression step were mechanical aptitude and the Block II examination score. These results also indicate that the administrative advanced measurement techniques were very successful predictors of final course average in the administrative courses.

Summary and Conclusions

One purpose of the present program was to describe, on an introductory basis, methods for providing a unifying core within the ISD technique as employed within the Air Force technical training context. A second purpose was to demonstrate the adaptability and utility of several advanced measurement and evaluation techniques to Air Force student measurement.

Multidimensional scaling and cluster analytic techniques were employed to order job analytic data for two Air Force technical specialties--electronics and administrative specialist. For both specialties, the results from both techniques were similar. Accordingly, it was held that either the multidimensional scaling or the clustering technique will provide the required definitional nexus within the ISD cycle.

A number of advanced measurement methods and techniques were developed on the basis of the job factors extracted from the job analytic data for each specialty.

Forty-one electronics students and 31 administrative students were administered the respective advanced measurement techniques. The students also completed a questionnaire regarding their attitudes toward some of the advanced measurement techniques. In addition, 10 instructors from each school were interviewed in order to assess their opinion of the new measurement methods.

The scores derived from the advanced measurement techniques were analyzed in regard to validity, reliability, uniqueness, and cost/effectiveness. Within each specialty, several of the advanced measurement techniques were indicated to possess psychometric, cost/benefit, and related properties which support a contention favoring their adoption. For the electronics specialty, these included: sequential testing, figural systems, confidence testing, and scoring on the basis of theory of signal detection (d'). For the administrative specialist, the supported advanced measurement techniques included: technical words, absurdity recognition, partial knowledge scoring, and d' scoring. At least for the two job specialties considered here, the following conclusions appear warranted:

1. The multidimensional scaling and the cluster analytic techniques appear to possess merit as methods for ordering job analytic data and for providing coherency within ISD application .

2. Extension of student achievement paper and pencil testing beyond the multiple choice format is warranted.

REFERENCES

Atkinson, J. Notes concerning the generality of the theory of achievement motivation. In J. Atkinson and T. Feather (Eds.), A theory of achievement motivation. New York: Wiley, 1966. Pp. 162-168.

Atkinson, J., & Feather, T. (Eds.) A theory of achievement motivation. New York: Wiley, 1966.

Atkinson, J., & O'Connor, P. Effects of ability grouping in schools related to individual differences in achievement-related motivation. Project 1283 of Cooperative Research Program of the Office of Education, United States Department of Health, Education, and Welfare.

Bergman, B. The effects of group size, personal space, and success-failure upon physiological arousal, test performance, and questionnaire response. Unpublished doctorial dissertation, Temple University, 1971.

Bergman, B., & Siegel, A. Training evaluation and student achievement measurement: A review of the literature. AFHRL-TR-72-3, AD-747 040. Lowry AFB, Colo.: Technical Training Division, Air Force Human Resources Laboratory, January 1972.

Bersoff, D., & Ericson, C. A precise and valid measure of behavior and behavior change. Proceedings of the annual convention of the American Psychological Association, 1971, 6(2), 555-556.

Dixon, W. (Ed.) BMD biomedical computer programs. Berkeley: University of Calif. Press, 1967.

Echternacht, G., Sellman, W., Boldt, R., & Young, J. An evaluation of the feasibility of confidence testing as a diagnostic aid in technical training. AFHRL-TR-71-33, AD-734 032. Lowry AFB, Colo.: Technical Training Division, Air Force Human Resources Laboratory, July 1971.

Elliott, P. Appendix 1 - tables of d'. In Swets, J. (Ed.), Signal detection and recognition by human observers. New York: Wiley, 1964.

Guilford, J., & Hoepfner, R. The analysis of intelligence. New York: McGraw-Hill, 1971.

Harrison, J., & MacKinnan, P. Central effect of epinephrine and norepinephrine on the palmar sweat index. American Journal of Physiology, 1963, 204, 785-788.

Harrison, J., & MacKinnan, P. Physiological role of the adrenal medulla in the palmar anhydratic response to stress. Journal of Applied Physiology, 1966, 21(1), 88-92.

Helm, C., Messick, S., & Tucker, L. Psychological models for relating discrimination and magnitude estimation scales. Psychological Review, 1961, 68, 167-177.

Instructional System Development. Washington, D. C.: Department of the Air Force, Air Force Manual 50-2, 1970.

Johnson, J., & Dobbs, J. Enumeration of active sweat glands: A simple physiological indicator of psychological changes. Nursing Research, 1967, 16(3), 273-276.

Johnson, S. Hierarchical clustering schemes. Psychometrika, 1967, 31(3), 241-254.

Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. Psychometrika, 1958, 23, 187-200.

Martens, R. Effects of an audience on learning and performance of a complex motor skill. Journal of Personality and Social Psychology, 1969, 12(3), 252-260.

Messick, S. An empirical evaluation of multidimensional successive intervals. Psychometrika, 1956, 21, 367-376.

Messick, S., & Abelson, R. The additive constant problem in multidimensional scaling. Psychometrika, 1956, 21, 1-17.

Morgan, J. The achievement motive and economic behavior. In J. Atkinson & T. Feather (Eds.), A theory of achievement motivation. New York: Wiley, 1966. Pp. 205-230.

Peterson, C., & Beach, L. Man as an intuitive statistician. Psychological Bulletin, 1967, 68, 29-46.

Schultz, D., & Siegel, A. Post-training performance criterion development and application: A multidimensional scaling analysis of the job performance of naval aviation electronics technicians. Wayne, Pa.: Applied Psychological Services, 1962.

Schultz, D., & Siegel, A. Post-training performance criterion development and application: A multidimensional scaling analysis of the circuit types repaired by naval aviation electronics technicians. Wayne, Pa.: Applied Psychological Services, 1963.

Shepard, R., Romney, A., & Nerlov, S. Multidimensional scaling: Theory and applications in the behavioral sciences. New York: Seminar Press, 1972.

Siegel, A., Bergman, B., Federman, P., & Sellman, W. Some techniques for the evaluation of technical training courses and students. AFHRL-TR-72-15, AD-753 094. Lowry AFB, Colo.: Technical Training Division, Air Force Human Resources Laboratory, February 1972.

Siegel, A., Fischl, M., & Pfeiffer, M. Personnel psychophysics: Terminal threshold and signal detection theoretic applications to performance assessment. Wayne, Pa.: Applied Psychological Services, 1968.

Siegel, A., & Schultz, D. Post-training performance criterion development and application: A comparative multidimensional scaling analysis of the tasks performed by naval aviation electronics technicians at two job levels. Wayne, Pa.: Applied Psychological Services, 1963.

Siegel, A., & Smith, R. A multidimensional scaling analysis of the job of the civil defense director. Wayne, Pa.: Applied Psychological Services, 1965.

Stodtbeck, F., McDonald, M., & Rosen, B. Evaluation of occupations: A reflection of Jewish and Italian mobility differences. American Sociological Review, 1957, 22, 546-553.

Sutarman & Thomson, A. A new technique for enumerating active sweat glands in man. Journal of Physiology (London), 1952, 117, 51-52.

Swets, J., Tanner, W., & Birdsall, T. Decision processes in perception. In J. Swets (Ed.), Signal detection and recognition by human observers. New York: Wiley, 1964.

Torgerson, W. Theory and methods of scaling. New York: Wiley, 1958.

Torgerson, W. Multidimensional scaling: I. Theory and method. Psychometrika, 1952, 17, 401-419.

APPENDICES

APPENDIX A

The First 12 Electronics Regression Equations Showing the Increase in Predictive Efficiency for Each Step in the Regression Process

One Step (R = .636)

$y' = 65.586 + .899x_5$

Two Steps (R = .669)

$y' = 41.255 + .800x_5 + .299x_{15}$

Three Steps (R = .693)

$y' = 37.943 + .304x_5 + .042x_7 + .330x_{15}$

Four Steps (R = .713)

$y' = 34.843 + .183x_5 + .042x_7 + .287x_{15} + .112x_{16}$

Five Steps (R = .734)

$y' = 34.516 + .060x_3 + .210x_5 + .036x_7 = .221x_{15} + .133x_{16}$

Six Steps (R = .751)

$y' = 41.679 + .067x_3 + .082x_5 + .047x_7 - .015x_9 + .167x_{15} + 120x_{16}$

Seven Steps (R = .762)

$y' = 40.337 + 1.051x_1 + .066x_3 - .018x_5 + .045x_7 - .018x_9 + .177x_{15} + .119x_{16}$

Eight Steps (R = .762)

$y' = 40.430 + 1.041x_1 + .066x_3 + .044x_7 - .018x_9 + .176x_{15} + .118x_{16}$

Nine Steps (R = .773)

$y' = 43.735 + 1.089x_1 + .074x_3 + .042x_7 - .023x_9 - .041x_{11} + .161x_{15} + .135x_{16}$

Ten Steps (R = .778)

$y' = 44.031 + 1.336x_1 + .081x_3 + .036x_7 - .023x_9 - .047x_{11} + .068x_{13} + .100x_{15} + .137x_{16}$

Eleven Steps (R = .783)

$y' = 48.022 + 1.393x_1 + .082x_3 + .034x_7 + .040x_8 - .025x_9 - .045x_{11} + .069x_{13} + .037x_{15} + 136x_{16}$

Twelve Steps (R = .786)

$y' = 49.070 + 1.500x_1 + .081x_3 + .032x_7 + .036x_8 - .026x_9 - .047x_{11} + .051x_{12} + .063x_{13} - .010x_{15} + .138x_{16}$

---

* The variable numbers are as follows: x_1 = sequential test level, x_2 = cognition of figural systems, x_3 = pictorial schematic reading, x_4 = pictorial absurdities, x_5 = conventional test score, x_6 = confidence test score, x_7 = signal detection score, x_8 = analogies, x_9 = need achievement (civilian), x_{10} = need achievement (Air Force), x_{11} = psychophysical arousal, x_{12} = general aptitude, x_{13} = mechanical aptitude, x_{14} = administrative aptitude, x_{15} = electronics aptitude, x_{16} = block IV exam score.

APPENDIX B

The First 12 Administrative Regression Equations Showing the Increase in Prediction of Final Course Average for Each Step in the Regression Process

One Step (R = .656)

$y' = 65.552 + .341x_{13}$

Two Steps (R = .769)

$y' = 51.117 + .997x_1 + .241x_{13}$

Three Steps (R = .824)

$y' = 48.457 + .778x_1 + .076x_4 + .193x_{13}$

Four Steps (R = .851)

$y' = 53.415 + .961x_1 + .081x_4 - .042x_6 + .195x_{13}$

Five Steps (R = .870)

$y' = 41.616 + .912x_1 + .069x_4 - .046x_6 + .150x_{13} + .202x_{16}$

Six Steps (R = .883)

$y' = 37.196 + .849x_1 + .053x_4 - .037x_6 + .064x_{11} + .139x_{13} + .227x_{16}$

Seven Steps (R = .893)

$y' = 24.948 + .769x_1 + .060x_4 - .047x_6 + .062x_8 + .070x_{11} + .157x_{13} + .215x_{16}$

Eight Steps (R = .899)

$y' = 25.870 + .715x_1 + .312x_2 + .070x_4 - .051x_6 + .061x_8 + .067x_{11} + .133x_{13} + .182x_{16}$

Nine Steps (R = .906)

$y' = 18.159 + .744x_1 + .442x_2 + .080x_4 - .052x_6 + .071x_8 + .031x_{10} + .062x_{11} + .084x_{13} + .191x_{16}$

Ten Steps (R = .912)

$y' = 12.154 + .674x_1 + .604x_2 - .610x_3 + .226x_4 - .041x_6 + .076x_8 + .052x_{10} + .071x_{11} + .078x_{13} + .201x_{16}$

Eleven Steps (R = .916)

$y' = 1.056 + .668x_1 + .705x_2 - .888x_3 + .290x_4 - .041x_6 + .107x_8 + .040x_9 + .040x_{10} + .062x_{11} + .039x_{13} + .224x_{16}$

Twelve Steps (R = .920)

$y' = -6.021 + .695x_1 + .750x_2 - .818x_3 + .276x_4 - .057x_6 + .051x_7 + .096x_8 + .043x_9 + .042x_{10} + .044x_{11} + .014x_{13} + .232x_{16}$

---

* The variable numbers are as follows: x_1 = technical words, x_2 = absurdities, x_3 = partial knowledge, x_4 = partial knowledge (d'), x_5 = typing I, x_6 = typing II, x_7 = typing III, x_8 = typing IV, x_9 = need achievement (civilian), x_{10} = need achievement (Air Force, x_{11} = psychophysiological arousal, x_{12} = general aptitude, x_{13} = mechanical aptitude, x_{14} = administrative aptitude, x_{15} = electronic aptitude, x_{16} = block exam score.

APPENDIX C

The First 11 Administrative Regression Equations Showing the Increase in Prediction of Course Hours for Each Step in the Regression Process

One Step (R = .677)

$y' = 266.967 - 7.870x_1$

Two Steps (R = .702)

$y' = 419.879 - 7.26x_1 - .620x_5$

Three Steps (R = .713)

$y' = 452.073 - 7.863x_1 - .763x_5 + .398x_{13}$

Four Steps (R = .741)

$y' = 535.339 - 7.276x_1 - .819x_5 + .707x_{13} - 1.108x_{16}$

Five Steps (R = .759)

$y' = 558.426 - 7.828x_1 - .746x_5 - .216x_9 + 1.094x_{13} - 1.348x_{16}$

Six Steps (R = .773)

$y' = 603.721 - 7.541x_1 - 2.179x_2 - .881x_5 - .270x_9 + 1.383x_{13} - 1.264x_{16}$

Seven Steps (R = .785)

$y' = 647.305 - 7.109x_1 - 2.708x_2 - 1.030x_5 - .298x_9 + 1.891x_{13} - .622x_{15} - 1.188x_{16}$

Eight Steps (R = .798)

$y' = 742.568 - 6.712x_1 - 2.832x_2 - .986x_5 - .389x_8 - .426x_9 + 2.065x_{13} - .747x_{15} - 1.225x_{16}$

Nine Steps (R = .807)

$y' = 822.692 - 6.941x_1 - 3.390x_2 - 1.231x_5 + .150x_6 - .513x_8 - .448x_9 + 2.259x_{13} - .869x_{15} - 1.297x_{16}$

Ten Steps (R = .812)

$y' = 770.384 - 6.860x_1 - 2.822x_2 - 1.045x_5 + .135x_6 - .553x_8 - .554x_9 + .173x_{10} + 2.058x_{13} - .775x_{15} - 1.257x_{16}$

Eleven Steps (R = .813)

$y' = 790.625 - 6.837x_1 - 2.871x_2 - 1.058x_5 + .140x_6 - .600x_8 - .588x_9 + .172x_{10} - .170x_{12} + 2.203x_{13} - .752x_{15} - 1.259x_{16}$

APPENDIX D

The First 12 Administrative Regression Equations Showing the Increase in Prediction of Final Course Average Divided by the Square Root of Course Hours for Each Step in the Regression Process

One Step (R = .735)

$y' = 20.342 + 3.185x_1$

Two Steps (R = .777)

$y' = -9.721 + 2.721x_1 + .465x_{16}$

Three Steps (R = .801)

$y' = -15.548 + 2.535x_1 + .146x_{11} + .467x_{16}$

Four Steps (R = .820)

$y' = -66.890 + 2.342x_1 + .214x_5 + .153x_{11} + .444x_{16}$

Five Steps (R = .831)

$y' = -76.619 + 2.559x_1 + .277x_5 - .056x_6 + .129x_{11} + .472x_{16}$

Six Steps (R = .842)

$y' = -84.430 + 2.403x_1 + .714x_2 + .321x_5 - .067x_6 + .130x_{11} + .386x_{16}$

Seven Steps (R = .849)

$y' = -91.569 + 2.495x_1 + .850x_2 + .342x_5 - .071x_6 + .134x_{11} - .132x_{12} + .453x_{16}$

Eight Steps (R = .852)

$y' = -95.736 + 2.524x_1 + .910x_2 + .339x_5 - .069x_6 + .033x_{10} + .127x_{11} - .141x_{12} + .451x_{16}$

Nine Steps (R = .856)

$y' = -102.033 + 2.486x_1 + .885x_2 + .352x_5 - .064x_6 + .044x_{10} + .140x_{11} - .245x_{12} + .180x_{14} + .398x_{16}$

Ten Steps (R = .858)

$y' = -130.780 + 2.526x_1 + 1.104x_2 + .426x_5 - .067x_6 + .079x_{10} + .154x_{11} - .228x_{12} - .178x_{13} + .280x_{14} + .419x_{16}$

Eleven Steps (R = .861)

$y' = -162.186 + 2.535x_1 + 1.274x_2 + .475x_5 - .101x_6 + .093x_7 + .096x_{10} + .132x_{11} - .197x_{12} - .256x_{13} + .271x_{14} + .447x_{16}$

Twelve Steps (R = .863)

$y' = -171.271 + 2.487x_1 + 1.263x_2 + .481x_5 - .103x_6 + .077x_7 + .046x_8 + .100x_{10} + .143x_{11} - .182x_{12} - .256x_{13} + .272x_{14} + .443x_{16}$

APPENDIX E

ELECTRONICS PRINCIPLES QUESTIONNAIRE

Name (print)	Date

Directions

Please indicate the extent to which you agree or disagree with each of the statements included in this booklet. The statements refer to the tests you have taken over the past few days. Indicate whether you: strongly agree (SA), agree (A), mildly agree (MA), mildly disagree (MD), disagree (D), or strongly disagree (SD) with each statement by circling the option which most closely reflects your true feeling. Please do not leave any statements unanswered.

CIRCLE YOUR ANSWER

1. The multiple choice test which allows one to assign confidence levels is fairer to the student than traditional multiple choice tests. (Circle one)	SA	A	MA	MD	D	SD
2. Traditional multiple choice questions test the electronic/electrical concepts of this course better than the kind of questions where you had to complete an analogy.	SA	A	MA	MD	D	SD
3. Tests like the component identification test are more relevant to electronic/electrical work than most multiple choice tests. (Circle one)	SA	A	MA	MD	D	SD
4. The tests that we took the past few days were more interesting to take than the usual tests in this school. (Circle one)	SA	A	MA	MD	D	SD
5. I prefer multiple choice test questions over the kind of questions in the absurdities test. (Circle one)	SA	A	MA	MD	D	SD

6.	I would rather take a plain multiple choice test than a test in which I must assign confidence levels to my answers. (Circle one)	SA	A	MA	MD	D	SD
7.	I would prefer to take more analogies tests and fewer of the usual type of school tests. (Circle one)	SA	A	MA	MD	D	SD
8.	I would like to avoid taking any more component identification tests in the future. (Circle one)	SA	A	MA	MD	D	SD
9.	I did not like the tests we took the past few days because the items were too difficult. (Circle one)	SA	A	MA	MD	D	SD
10.	The test in which we had to draw lines connecting various components and parts is preferable to most multiple choice tests. (Circle one)	SA	A	MA	MD	D	SD
11.	I did not like the tests we took the past few days because I could not understand the directions. (Circle one)	SA	A	MA	MD	D	SD
12.	I did not like the tests we took the past few days because they interfered with my time schedule and studies. (Circle one)	SA	A	MA	MD	D	SD
13.	I would never again like to take a test like the one in which we had to draw lines connecting various components and parts. (Circle one)	SA	A	MA	MD	D	SD
14.	I wouldn't mind taking part in any future testing programs like the one we had the past few days. (Circle one)	SA	A	MA	MD	D	SD
15.	I did not like the tests we took the past few days because the questions were too complicated. (Circle one)	SA	A	MA	MD	D	SD
16.	The tests we took the past few days seem to measure what was taught in the course. (Circle one)	SA	A	MA	MD	D	SD

APPENDIX F

ADMINISTRATIVE SPECIALIST QUESTIONNAIRE

______________________________ ____________________

Name (print) Date

Directions

Please indicate the extent to which you agree or disagree with each of the following statements included in this booklet. The statements refer to the tests you have taken over the past few days. Indicate whether you strongly agree (SA), agree (A), mildly agree (MA), mildly disagree (MD), disagree (D), or strongly disagree (SD) with each statement by circling the option which most closely reflects your true feeling. Please do not leave any statements unanswered.

CIRCLE YOUR ANSWER

1. I prefer a multiple choice test in which I eliminate the wrong answers rather than pick the right answers. (Circle one)	SA	A	MA	MD	D	SD
2. I prefer a multiple choice test over a test like the absurdities test. (Circle one)	SA	A	MA	MD	D	SD
3. I think tests like the technical words test are relevant to working as an administrative specialist. (Circle one)	SA	A	MA	MD	D	SD
4. I would prefer to take a usual multiple choice test over a test in which I have to eliminate all the wrong answers. (Circle one)	SA	A	MA	MD	D	SD
5. I think the absurdities test was most interesting and preferable to multiple choice tests. (Circle one)	SA	A	MA	MD	D	SD
6. I don't think the technical words test was very useful and I would not like to take any more tests like that in the future. (Circle one)	SA	A	MA	MD	D	SD

7. The tests that we took the past few days were more interesting than the usual school tests. (Circle one)	SA	A	MA	MD	D	SD
8. I did not like the tests we took the past few days because the items were too difficult. (Circle one)	SA	A	MA	MD	D	SD
9. I did not like the tests we took the past few days because I could not understand the directions. (Circle one)	SA	A	MA	MD	D	SD
10. I did not like the tests we took the past few days because they interfered with my time schedule and studies. (Circle one)	SA	A	MA	MD	D	SD
11. I wouldn't mind taking part in any future testing programs like the one we had the past few days. (Circle one)	SA	A	MA	MD	D	SD
12. I did not like the tests we took the past few days because they didn't measure what is taught in the course. (Circle one)	SA	A	MA	MD	D	SD