AD A057507

LEVEL II

12

ARI TECHNICAL REPORT
TR-78-A13

Development and Evaluation of a Videotape Simulation Performance Test

by

Dan D. Jennings, Jr., Anthony W. Kendall,
and Myron A. Robinson

DATA-DESIGN Laboratories
7925 Center Avenue
Cucamonga, California 91730

DDC
AUG 15 1978
F

MAY 1978

Contract DAHC-19-76-C-0032

Prepared for



U.S. ARMY RESEARCH INSTITUTE
for the BEHAVIORAL and SOCIAL SCIENCES
5001 Eisenhower Avenue
Alexandria, Virginia 22333

78 08 15 083

Approved for public release; distribution unlimited.

AU NO.
DDC FILE COPY

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

A Field Operating Agency under the Jurisdiction of the Deputy Chief of Staff for Personnel

JOSEPH ZEIDNER
Acting Technical Director

W. C. MAUS
COL, GS
Commander

Research accomplished
under contract to the Department of the Army

Data-Design Laboratories

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U. S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-P, 5001 Eisenhower Avenue, Alexandria, Virginia 22333.

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

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

(18) ARI

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

(19) 1. REPORT NUMBER: TR-78-A13

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

(6) 4. TITLE (and Subtitle): DEVELOPMENT AND EVALUATION OF A VIDEOTAPE SIMULATION PERFORMANCE TEST.

5. TYPE OF REPORT & PERIOD COVERED: Final Report for Period 1 May 1976 to 31 July 1977

(14) DDL

6. PERFORMING ORG. REPORT NUMBER: TSD-78-2001

(10) 7. AUTHOR(s): Dan D. Jennings, Jr.
Anthony W. Kendall
Myron A. Robinson

(15) 8. CONTRACT OR GRANT NUMBER(s): DAHC19-76-C-0032

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Data-Design Laboratories
7925 Center Avenue
Cucamonga, California 91730

(16) 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 2Q263731A770

(11) 11. CONTROLLING OFFICE NAME AND ADDRESS:
TRADOC, Army Training Support Center
Fort Eustis, Virginia

12. REPORT DATE: May 1978

13. NUMBER OF PAGES: 142

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):
U.S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, Virginia 22333

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

(12) 136p.

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

18. SUPPLEMENTARY NOTES

Research monitored technically by Milton H. Maier, Individual Training and Skill Evaluation Technical Area, ARI.

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

Simulation testing, synthetic testing, audio-visual simulation test, television simulation testing, videotape simulation testing, simulation performance testing.

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

A study was made to determine the feasibility of using videotape as the presentation mode in a simulation performance test of certain tasks in the Army's Carpentry and Masonry MOS. Two sets of procedures -- (1) task selection procedures, and (2) simulation procedures -- were developed which provide detailed guidance to future simulation test developers. A prototype simulation test was developed and validated against general performance ratings and a similar written instrument. The following was concluded: (1) application of the procedures enabled the selection of the more appropriate tasks

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

conclusions were made;

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

389081

LB

and task elements from a specified field of tasks but required a greater expenditure of human resources than may be typically resident in Army test development activities; (2) the fundamental question of the applicability of audio-visual simulation to test perceptual content was not conclusively answered; and (3) the use of television strictly for testing the perceptual content of lower skill level motor tasks such as those within the Carpentry and Masonry MOS appears somewhat limited; there appears, however, to be a decided favorable attitudinal bias, on the part of the test taker, towards television testing.

A companion volume was produced for ARI, "Guidelines for the Developers of Videotape Simulation Performance tests," P-78-1, which provides detailed guidance for test developers.

ACCESSION for
NTIS White Section ☑
DDC Buff Section ☐
UNANNOUNCED ☐
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
Dist. SPECIAL
A

SUMMARY AND CONCLUSIONS

BACKGROUND

The Army Research Institute (ARI) has, for several years, been exploring various methods of "synthetic performance testing" (Osborn, 1970) in an effort to identify alternatives to the full performance test, which is seen as more valid, but not as feasible, and the written test, which is seen as more feasible, but not generally as valid. The development of valid and reliable synthetic performance tests has significant impact on the performance-based criterion referenced Skill Qualification Tests (SQT) which form the evaluative heart of the Army's Enlisted Personnel Management System. One form of synthetic test which appeared promising was audio-visual simulation testing with television. Specifically, ARI was interested in assessing television simulation as a means of presenting perceptual and perceptual related psychomotor tasks within the Army's Carpentry and Masonry Military Occupational Specialty. It is this assessment which is addressed in this report.

PROCEDURES

The research design entailed the completion of four tasks: (1) a prototype task selection procedure was developed which ranked a given field of tasks on the basis of their common critical elements in order to permit the early analysis of those tasks deemed most appropriate for inclusion in a simulation test, (2) a simulation procedure was constructed which took tasks from the task procedure and provided a structured method for analyzing the common, critical elements in terms of their perceptual components and their feasibility for simulation testing, (3) a prototype simulation (television) test was constructed in conformity with the task selection and simulation procedures, and (4) the prototype test was evaluated in terms of its validity and feasibility when compared with performance and written tests.

Achievement of a second contract objective, the construction of guidelines for the developers of similar tests, involved a separate deliverable item and is not addressed in this report.

I 78 08 15 073

RESULTS

Both the task selection and the simulation procedures provided task element data and guidance generally as expected. The prototype simulation test, when correlated with less than ideal criterion measures, was shown to be significant but not necessarily of more validity than a similar, written test.

CONCLUSIONS

TASK SELECTION AND SIMULATION PROCEDURES

The application of the procedures enabled the selection of the more appropriate tasks and task components from a specified field of tasks critical to MOS 51A and 51B.

Use of the simulation procedures requires a greater expenditure of human resources than may typically be present in a test development agency.

APPLICABILITY OF A/V SIMULATION

The fundamental question of the applicability and validity of A/V simulation to test perceptual content was not conclusively answered because of a number of problems discussed in the text of the report.

The use of television as a simulation means, strictly for testing the perceptual content of lower skill level motor tasks such as those within the carpentry and masonry MOS appears somewhat limited; there appears, however, to be a decided favorable attitudinal bias, on the part of the test takers, towards television testing.

TABLE OF CONTENTS

	Page
Background	7
Requirement	7
Statement of the Problem	7
Limitations on Performance of the Research	10
Purpose	11
Development and Application of the Task Selection Procedures	11
Development and Application of the Simulation Procedures	13
Scope and Limitations of the Simulation Procedures	20
Production and Evaluation of the Simulation Test	26
Production of the Test	26
Identification of External Criteria	27
Selection of Matched Groups	29
Test Administration Procedures	30
Pooling of Test Data	30
Performance on Course Modules	32
Performance Data on Written Tests	33
Performance Data on the Simulation Test	33
Unit and Overall Test Validity	33
Assessment of Attitudes Toward Simulation	34
Structured Items	35
Open-end Items	38
Comparative Costs of Videotape Simulation Tests	39
Lessons Learned	44
Limitations Affecting the Validation of the Simulation Test	44
Problems Inherent in Present Application of A/V Simulation Tests	45
Perceptual Content in Job	46
Similarity of Test and Job Response	46
General Considerations Concerning the Use of Television Simulation	46
Motivational/Attitudinal	46
Resolution/Acuity	47
Costs of A/V Simulation Versus Written and Performance Testing	48
Simulation of Dynamic Tasks	48
Conclusions	50
Task Selection Simulation Procedures	50
Applicability of A/V Simulation	50
References	51

TABLE OF CONTENTS (Continued)

Appendix A--Task Selection Procedures
Appendix B--Simulation Procedures
Appendix C--Simulation Test Audio Script and Answer Sheet
Appendix D--Written Test
Appendix E--Questionnaire

LIST OF TABLES

Table		Page
1	Frequency of Steps in Job Sample Tests Categorized According to Knowledge and Skill Requirements	9
2	Element Analysis to Identify Components and Stimulus Variables	19
3	Element Analysis to Determine the Importance of Stimulus Variables Associated with a Critical Element	21
4	Element Analysis for Preliminary Determination of Test Mode	22
5	Element Analysis to Determine Possible Test Response	23
6	Relation of Simulation Test Items to Critical Elements and Stimulus Variables	24
7	Relation of External Criteria to Test Items	28
8	Description of Experimental Groups and Criteria	29
9	Analysis of Variance for Simulation (Television) Test Scores	31
10	Analysis of Variance for Written Test Scores	31
11	Means and Standard Deviations of Rating Scores on Criterion Modules by Associated Simulation Test Unit and Group	32
12	Means and Standard Deviations of Percentages of Correct Responses on the Written Test for Carpenters and Utility Workers	33
13	Means and Standard Deviations of Percentages of Correct Responses on the Simulation Test for Carpenters and Utility Workers	34
14	Correlation of Simulation and Written Test Scores With Ratings on Criterion Modules	35
15	Scaled Questionnaire Data	36
16	Development and Validation Expenditures	41
17	Videotaped Test Material Cost (Development)	42
18	Videotaped Production Equipment Configuration	44
A-1	Job Task Summary Sheet	61
B-1	Element Analysis (A)	77
B-2	Element Analysis (B)	79
B-3	Element Analysis (C)	84

LIST OF TABLES

Table		Page
B-4	Determining the Importance of Each Stimulus Variable Associated With a Critical Element	86
B-5	Determining the Appropriateness of an Audio-Visual Presentation Mode for a Simulated Skill Qualification Test	87
B-6	Comparison of Response Requirements to Stimulus Variable	91
B-7	Evaluation of the Job-Functional-Context	102

LIST OF ILLUSTRATIONS

Figure		Page
1	Task Selection Matrix	15
2	Job Task Summary Sheet	17
A-1	Task Selection Matrix	65
A-2	Task Selection Procedures	67
B-1	Simulation Algorithm	73
B-2	Preliminary Test Mode Selection	82
B-3	Test Realism	89
B-4	Response Realism	97
B-5	Test Mode Selection	100
B-6	Representation of Job Context	104
B-7	Final Assessment of Presentation Realism	106
B-8	Test Development	115
D-1	Wall Form	136

BACKGROUND

REQUIREMENT

Under contract DAHC 19-76-C-0032, Data-Design's efforts were directed towards the accomplishment of two basic objectives: (1) develop and evaluate the methodology necessary to construct an audio-visual simulation (television) performance test of tasks within the Army's Carpentry and Masonry Military Occupational Specialties (MOS), and (2) produce the guidelines which will enable Army test developers to construct simulation performance tests for similar MOS.

Achievement of the first objective required the performance of four tasks: (1) the development of a prototype task procedure to analyze and classify perceptual and perceptual related psychomotor tasks according to their common, critical elements, (2) the development of a simulation procedure which will take tasks from the task model and provide a structured method for analyzing the common, critical elements in terms of their feasibility for simulation testing, (3) the development of a prototype simulation test, and (4) the evaluation of the prototype test, in terms of its validity and costs when compared with performance and written tests.

Achievement of the second objective, production of the guidelines, involved a separate deliverable item under the contract and will not be addressed in this report.

STATEMENT OF THE PROBLEM

The Army's Enlisted Personnel Management System (EPMS) is, to a great extent, dependent upon the construction and utilization of Skill Qualification Tests (SQT) which are fair, valid, and reliable measurements of present performance and predictors of future performance. The procedures and requirements for constructing these tests are specified by official directive (ITED, 1976). To be useful, SQT must be administered feasibly in the field, under a variety of conditions, to the many thousands of soldiers worldwide. SQT are performance based, criterion referenced test instruments, which are to be constructed as hands-on, full performance tests whenever they can be

PRECEDING PAGE BLANK

administered feasibly as such. Feasibility is used here in the sense of the cost-effective use of limited resources. For many tasks, full performance tests involve the dedication of costly equipment and manpower each time the test is given. Raters must be trained, test sites and conditions must (to the extent practicable) be standardized, and in many cases, key personnel are assigned as raters and are therefore lost to their primary duty stations during test administration. When faced with tasks which simply cannot be administered feasibly in the hands-on components, or if the component is full, the test developer has two alternatives: (1) the performance certification component, or (2) the written component.

Evidence has shown, however, that paper-and-pencil tests are of relatively low validity when correlated with performance tests. Pickering and Anderson (1976) conclude that the correlations between performance and theory tests and performance and knowledge tests "are not high enough to justify the substitution of job knowledge tests for job performance tests." Foley (1974), presented correlations of job performance tests with theory tests and job knowledge tests and the ranges (.03, .36) for performance and theory, and (.10, .55) for performance and knowledge, indicate a low validity for these written instruments.

Higher correlations between job knowledge and job performance (work sample) tests were reported by Vineberg and Taylor (1972), in their comparison of criterion instruments in four Army jobs (Armor Crewman, Repairman, Supply Specialist, and Cook). These correlations (with the effects of time on the job partialled out) ranged from .49 for the Armor Crewman and the Repairman to .65 for the Supply Specialist. It should be pointed out, however, that the four jobs were categorized in terms of their knowledge and skill requirements as shown in Table 1. It can readily be seen that the high correlation for the Supply Specialist is due, at least in part, to the absence of skill requirements for the job.

Indeed, the authors indicate that this relationship is of particular significance for their analysis in that the job of the Supply Specialist, "represents one of the purest examples of a job where knowledge rather than skill is sufficient to support performance" (p.20).

Army test developers are faced with essentially two unsatisfying choices; (1), a performance test which is more valid but typically not as feasible, or

TABLE 1. FREQUENCY OF STEPS IN JOB SAMPLE TESTS CATEGORIZED ACCORDING TO KNOWLEDGE AND SKILL REQUIREMENTS

Requirements	Armor Crewman	Repairman	Supply Specialist	Cook
Knowledge Alone	338	165	153	145
Cognitive Skill and Knowledge	0	4	3	11
Perceptual-Motor Skill and Knowledge	16	6	0	2
Total Number of Steps	354	175	156	158

NOTE: From Vineberg and Taylor (1972)

(2), a written test which is typically more feasible but not as valid. Consequently there was a desire to develop a broader range of alternatives. The Army Research Institute (ARI) began exploring various methods of "synthetic performance testing" (Osborn, 1970) in order to develop alternatives between the extremes of performance and written tests. One form of synthetic testing which appeared promising was audio-visual simulation as a means of presenting perceptual or perceptual related psychomotor tasks.

ARI designated skill levels 1 and 2 of MOS 51B (Carpentry and Masonry Specialist) as an appropriate vehicle for assessing the effectiveness of SQT testing via audio-visual simulation. These skill levels are nonsupervisory. Tasks such as "place and finish concrete," and "construct and erect wall forms" are extremely poor candidates for both performance and written tests. The expense of administering full performance tests of these items to a large, worldwide population would be significant and likely prohibitive, and the validity of a written test appears questionable. Since both of these tasks appeared to include a substantial perceptual component (i.e., fine perceptual discriminations must precede motor behavior), it appeared that audio-visual simulation was viewed as potentially an acceptable alternative.

A preliminary analysis of skill levels 1 and 2 for the Carpentry and Masonry MOS indicated that the information was primarily equipment and

procedure oriented. For example, fabrication of a frame is typically an ongoing process which involves the response to stimuli most frequently given in the form of visual or spatial cues.

The research team proposed that videotape would be the most suitable equipment to convey subject matter which is both visual and spatial in nature. In comparison with the other equipments currently available at each active Army battalion, videotape appeared to offer the most advantages in terms of production features and message impact. Print messages have a low sensory impact. Essentially, they lack the capacity to attract busy individuals to their message. Audiotape is monosensory; that is to say it lacks the visual element entirely and was inappropriate. Slides or filmstrips lack the dimension of motion, and by this deficiency appeared to exclude any real ability to reproduce the attributes of an ongoing procedure. Of the equipments available at the battalion sites, only the Beseler Cue/See devices and videotape had the ability to present subjects which required motion, sound or color, and there appeared to be certain advantages to videotape production that made a better choice.

Working with film can be a very time consuming venture. Once all the footage is shot, the film is mailed away to be developed, and when returned it is reviewed and edited. If there are any problems with the film, or additional scenes are needed, the whole cycle has to be repeated. In working with videotape, there is no time delay between shooting material and reviewing it. Immediately after the footage is shot, it can be reviewed on-the-spot with a portable monitor. Then, if additional shots are required, they can be taken at that time.

ARI concurred with the proposed test presentation mode and the subsequent research effort reflected that decision.

LIMITATIONS ON PERFORMANCE OF THE RESEARCH

When this research was planned, it was anticipated that MOS 51A (Utility worker) and 51B (Carpenter) would be combined to form a new MOS 51B (Carpenter and Mason). This combination would have resulted in the construction of a Skill Qualification Test (SQT) for the MOS. The written and hands-on components of the SQT would then have formed the evaluative criteria for the subsequently developed simulation test. However, the new

MOS was not created, and as a consequence, no SQT was developed for either 51A or 51B. In the absence of an SQT, it was necessary to obtain criterion measures elsewhere. The most appropriate alternative source available was the performance ratings from certain module tests given to trainees during the advanced individual training of both MOS 51A10 and 51B10. Thus, criteria for the validation of the prototype simulation test were not item specific, but generalized from the module test ratings.

PURPOSE

The specific purpose of this research was to evaluate a televised simulation test of a few tasks of an SQT, as a medium for testing perceptual and perceptual related components within MOS 51B. This research is part of a broader ARI effort to assess the overall effectiveness of an audio-visual simulation SQT.

This evaluation investigated both the stimulus and response dimensions of these components and the degree to which they might be faithfully replicated by television. It also included an evaluation of the feasibility of developing and administering television simulation tests in the field. Feasibility is measured here by two criteria: (1) ability to develop the tests using the manpower typically resident in Army test development activities, and (2) administration costs which are significantly less than those incurred in the administration of full performance tests of the same tasks.

DEVELOPMENT AND APPLICATION OF THE TASK SELECTION PROCEDURES

Development

The Army test developer (today) can be faced with a list of critical tasks ranging from as few as 30 to more than 250, depending upon the MOS. Current guidance (Osborn, et al, 1977) specifies that no more than 76 tasks may be included in any one SQT. Because time is a resource typically in short supply at test development activities, an efficient, workable methodology for reducing the task field to a manageable number is clearly desired. The Task Selection Procedures (Appendix A) were designed to aid the test developer in the systematic reduction and ordering of a given field of tasks. The

procedures allow assessment of each task, at the task element level, for its commonality and criticality when compared to the entire field of tasks in that MOS at a given skill level. The resultant task list is rank ordered from the most common and critical task to the least.

It was felt that an existing taxonomy of general verbs might be identified which could be employed in developing the procedures. Building upon the pioneering work of Cotterman (1959), Berliner, et al (1964) classified tasks into four major categories; i.e., perceptual, mediational, communicational, and motor processes. Bennett (1971) factor analyzed student judgments of task-related verbs and found four categories of task variables which he labeled as cognitive, social, procedural, and physical. Other task taxonomies were developed by Gagne (1962), Folley (1964), and Fleishman and his associates (summarized in 1975) among others.

Is it possible to construct a universal task classification model? Fleishman (1975) states: "The search for a single general taxonomy is not likely to be successful for all purposes. We may, indeed, need several task classification systems for several purposes, with the linkage between them understood and specified . . . Taxonomies are not out there to be discovered; some invention is required." (p. 1147).

Discussions with senior personnel at the SQT Branch, U.S. Army Engineer School, Fort Belvoir, VA, and at Engineer training brigades at Fort Leonard Wood, MO, led to the belief that any single list of verbs ran the risk of being too general. This became especially evident upon studying the task descriptions of MOS 51A. Specialized verbs such as screed, trowel, and float are unique or have definitions which are unique to tasks dealing with concrete. Obviously, the difference between "floating" concrete (part of the finishing process) and floating a river (as in tactical amphibious operation) is significant and to class them as common behaviors on the basis of their common verbs would be misleading.

Procedures were developed, then, to allow individual test development organizations to utilize simple taxonomies in which the verbs are specific to each MOS and skill level. The user is directed to construct a matrix in which critical tasks are listed horizontally and the key verbs are listed vertically. The matrix allows for a heavier weighting of verbs or tasks which are considered by subject matter experts to be highly critical. The verbs listed in the vertical column can be taken from the performance steps as listed in

the Task Data Cards, Job Task Summary Sheets, or whatever job/task analysis form is in use at the particular organization.

The output of these procedures is a rank ordering of tasks in terms of criticality and commonality at the task element level.

Application

An inventory of tasks, designated as critical by the Army Engineer School's Task Analysis Branch, was obtained and plotted in the selection matrix as shown in Figure 1. Figure 2, the Job Task Summary Sheet, is typical of the task statements used as input data.

In order to demonstrate the reliability of the selection procedures, the same list of critical tasks was given to a senior noncommissioned officer in the test development activity at the Army Engineer School, Fort Belvoir, VA. The NCO applied the procedures to the list and with one exception, ranked the tasks in the same order as the first application. The difference was caused by the NCO's plotting the behavior designator "maintain tools" as the most common in the entire list. He reasoned properly, that the behavior of maintaining tools was an implied step in the performance of every task in the list, and plotted it as such. This points up a major limitation to the matrix; i.e., its dependence upon detailed task statements as input data. Strictly interpreted, the procedures allow the user to ignore any behaviors not explicit in the task statements.

Another NCO was asked to use the procedures to order a list of tasks from the MOS in which he was a subject matter expert; i.e., MOS 62F (Crane Operator). The procedures are generalizable to the extent that he was able to complete the matrix and rank order a sample list of 20 tasks, using JTSSs as input.

DEVELOPMENT AND APPLICATION OF THE SIMULATION PROCEDURES

Development

These procedures provide the test developer with a means of selecting appropriate perceptual content for A/V simulation tests and for developing

valid criterion referenced simulation tests. The following features are provided:

(1) An output of a simulation test that can be presented in a standard situation to examinees at many sites.

(2) A procedure for selecting only critical components of tasks.

(3) A procedure for presenting items in the job functional context.

(4) A procedure for sequencing activities as they will occur on the job.

(5) A procedure for identifying relevant stimuli for decision making.

(6) A procedure for identifying appropriate test responses to test stimuli.

(7) Procedures which can be applied to A/V simulation test construction with task data from other MOSs.

Development of the procedures was largely concerned with the question of how to construct a criterion-referenced test; a subject that has received considerable attention in recent years (Laabs, Main, Abrams, Steinnemann, 1975; Osborn, 1973; Panitz and Olivo, 1970; Swezey and Pearlstein, 1974; Osborn, Campbell, Ford, Hirshfeld and Maier, 1977). Two major differences between the simulation procedures and those strictly concerned with job sample, criterion-referenced tests are that: (1) these procedures are also concerned with assuring appropriate simulation of test items and (2) the procedures assume that each task element is associated with a behavioral objective which specifies that the examinee should be able to perform the task element under job conditions at a mastery level. (This is stated as an assumption because both sets of procedures start with the task element as a given and can only assume that it was properly derived in a task analysis model.)

The simulation procedures are presented in the format of an algorithm which leads the test developer through the following sequence of events:

(1) Analysis of the selected critical task elements to identify perceptual, cognitive, and motor components.[1]

(2) Selection of the perceptual components for videotaped simulation and assignment of motor and cognitive components to performance and written tests.

[1] An informal review of the simulation procedures by test development personnel at the Army Engineer School resulted in the terms cognitive and motor being redesignated as decision and action. This was done to enhance user acceptability and is reflected in subsequent element analysis tables.

BEHAVIOR DESIGNATORS	CRITICAL TASKS																
	Maintain carpenter tools and equipment.	Identify construction material by type and size.	Cut and install batter boards.	Construct and replace footers and columns.	Fabricate and install girders, floor joists and bridging.	Install and/or replace subfloor.	Frame walls and partitions.	Assemble roof trusses using template.	Install roof trusses.	Construct building roofs.	Assist in the fabrication and installation of forms for concrete footers, foundations, etc.	Install anchor bolts in concrete.	Construct joints in concrete	Prepare timber piles for driving.	Assist in the construction of pile bents for bridges, piers and wharves.	Place and finish concrete.	Assist in the fabrication and installation of forms for walls, columns, stairs and
Assemble																	X
Bore												X			X		
Maintain (Tools)	X																
Cut (Concrete)													X				
Cut (Wood)			(X)	(X)	(X)	(X)	(X)	(X)		(X)	(X)			(X)			(X)
Consolidate (Vib. Conc)																X	
Check (Inspect)				(X)	(X)		(X)	(X)		(X)	(X)	(X)		(X)			
Edge (Concrete)				X												X	
Emplace			X														
Excavate				X													
Finish (Concrete)				(X)*												(X)*	
Identify		X															
Interpret (Dwgs)						X											
Laminate					X												
Lay (Brick, Block)				X													
Lay (Sheathing, Shingles)										X							
Level						X											
Lower (w/jacks)				X													
Measure			(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)		(X)			(X)
Mix				X													
Nail			(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)							(X)
Oil/Wet (Forms & Bolts)				(X)							(X)	(X)	(X)				(X)
Operate (Chain Saw)														X	X		
Place (Concrete)				(X)*												(X)*	
Position							(X)	(X)	(X)			(X)					
Plumb				(X)			(X)		(X)	(X)							
Raise (w/jacks)				X													
Sharpen (Tools)	X																
Shape (Piles)														X			
Square							(X)	(X)		(X)				(X)			
Sharpen (Posts)			X														
Bolt															X		
	0	0	3	8	4	3	7	6	4	6	4	4	1	4	0	2	4

* Critical, though not common

1

CRITICAL TASKS											
Frame walls and partitions.	Assemble roof trusses using template.	Install roof trusses.	Construct building roofs.	Assist in the fabrication and installation of forms for concrete footers, foundations, etc.	Install anchor bolts in concrete.	Construct joints in concrete	Prepare timber piles for driving.	Assist in the construction of pile bents for bridges, piers and wharves.	Place and finish concrete.	Assist in the fabrication and installation of forms for walls, columns, stairs and floor slabs.	
										X	1
					X			X			2
											1
						X					1
(X)	(X)		(X)	(X)			(X)			(X)	10
									X		1
(X)	(X)		(X)	(X)	(X)		(X)				8
									X		2
											1
											1
									(X)*		2
											1
											1
											1
											1
			X								1
											1
											1
(X)	(X)	(X)	(X)	(X)	(X)		(X)			(X)	12
											1
(X)	(X)	(X)	(X)							(X)	9
				(X)	(X)	(X)				(X)	5
							X	X			2
									(X)*		2
(X)	(X)	(X)			(X)						4
(X)		(X)	(X)								4
											1
											1
							X				1
(X)	(X)		(X)				(X)				4
											1
								X			1
7	6	4	6	4	4	1	4	0	2	4	85 ÷ 32 = 2.7

TASKS, RANK ORDERED

1. Construct and replace footers and columns.
2. *Frame walls and partitions.*
3. Assemble roof trusses using template.
   Construct building roofs.
4. Fabricate and install girders, floor joists and bridging.
   Install roof trusses.
   Assist in the fabrication and installation of forms for concrete footers, foundations, etc.
   Install anchor bolts in concrete.
   Prepare timber piles for driving.
   Assist in the fabrication and installation of forms for walls, columns, stairs and floor slabs.
5. Cut and install batter boards.
   Install and/or replace subfloor.
6. Place and finish concrete.
7. Construct joints in concrete.
8. Maintain carpenter tools and equipment.
9. Identify construction material by type and size.
10. Assist in the construction of pile bents for bridges, piers, and wharves.

Figure 1. Task Selection Matrix

1		2
TASK: Place and finish concrete	**TASK CRITICALITY (CIRCLE ONE) C I N**	**ITEP TASK REFERENCE** _____ **ARTEP TASK REFERENCE** 14-35 **APPLICABLE MOS** EPMS 51B10 **SKILL LEVEL** 1 **%PERFORMING TASK** _____ **NO. PERFORMING TASK** _____ **FREQUENCY OF PERFORMANCE** _____ **JOB ANALYST** _____ **DATE PREPARED** 3 April 75
CONDITION(S): Given installed forms, mixed concrete, transporting equipment and masonry kit, an EPMS 51B10 as a concrete const. crew member under close supervision of EPMS 51B20 and/or 51H30.		
STANDARD(S): IAW const. prints, directions of supervisor and TM, concrete will be placed properly into forms and finished to prescribed finish.		
REFERENCES: TM 5-742 Concrete and Masonry June 70		

3 STEPS IN PERFORMANCE	4 STANDARD OF PERFORMANCE	5 MATERIALS, TOOLS, EQUIPMENT	6 SPECIFIC REFERENCE, ILLUSTRATION, OR SKETCH
1. Place plastic concrete into forms for slab.	Placement of concrete will start at far end of slab and each batch will be dumped against previously placed batch as directed by crew chief.	Installed forms, mixed concrete, wheel barrow or loading equipment, shovels, rakes.	TM 5-742, para 5-23, Fig. 5-13.
2. Place concrete into wall, beam, and girder forms.	Concrete will be placed in 6" to 24" lifts only, and maximum free fall of concrete will be 5' when placing concrete into wall forms. Concrete will be placed from each end and work to center of form when placing in beam or girder forms under close direction of crew chief.	Installed forms, mixed concrete wheel barrow or loading equipment, shovels, hoes, rakes.	TM 5-742, para 5-23, Fig. 5-13.
3. Consolidate concrete using mechanical vibrator.	Vibrator will be inserted into concrete at approximately 18" intervals for 5 to 15 seconds. Vibrator will be lowered using its own weight to penetrate through several inches. Vibrator will be used under close supervision of crew chief.	Mixed concrete, mechanical vibrator air compressor.	TM 5-742, para 5-24, Fig. 5-14.
4. Screed concrete to finish grade.	Screed board will be placed flat on wood or steel forms moving across placed concrete with a sawing action and forward motion. Screeding will be done twice over area to remove excess concrete brought up by first screeding.	Mixed concrete, screed board.	TM 5-742, para 5-27, Fig. 5-17.
5. Assist finishing concrete using wood float.	Wood float finishing will start at a time and in a manner as directed by crew chief. Wood float finishing will be used to remove high to low spots and imbed aggregate for other finishing operation.	Wood float placed concrete.	TM 5-742, para 5-27, Fig. 5-18.
6. Assist finishing concrete using steel trowel.	Steel troweling of concrete will start only as directed by crew chief. Steel troweling will produce a smooth even free of marks and ripples surface.	Steel trowel.	TM 5-742, para 7-27, Fig. 5-19.
7. Assist finishing concrete using broom.	Broomed finish will produce a rough non-skid surface. Procedures will be closely directed by crew chief.	Broom.	TM 5-742, para 7-27,
8. Edge concrete.	Edges of placed concrete will be edged to eliminate chipping of edges when removing forms. Edging procedures will be directed by crew chief.	Edger.	

Figure 2. Job Task Summary Sheet

(3) Identification of perceptual components in very specific stimulus terms to enable a determination of whether the A/V media can present the stimulus with adequate fidelity.

(4) Identification and analysis of the job and test taking response to all of the stimuli to be simulated.

(5) Identification of relevant stimuli in the job environment to insure the optimization of contextual cues in the simulation test.

(6) Assessment of the adequacy of the simulation mode in presenting all of the stimuli to be simulated.

(7) Sequencing test stimuli into a test format.

(8) Scripting the A/V test.

(9) Subject matter expert reviews of the test with revisions as necessary.

(10) Determination of test validity.

The algorithm format was selected as most appropriate for guiding Army test development personnel through a complex process. It is supplemented by narrative and operational examples of each step to enhance user acceptability. It also incorporates the use of current Army test development documentation.

Recognizing that the intended user was unlikely to be an expert in video-production, the procedures call for early participation of a media expert in the development process, and also reference an appended nontechnical discussion of the video-production process.

Application

The tasks identified as highly common and critical were then analyzed through the task selection procedures. As a result, three tasks: (1) maintain carpenter tools and equipment, (2) construct and erect concrete wall forms,[2/] and (3) place and finish concrete, were selected for inclusion in the simulation test.

[2/] During the first subject matter expert review of the candidate tasks, personnel at the U.S. Army Training Center (Engineer), Fort Leonard Wood, Mo., pointed out that the task "Construct and replace footers and columns," was rarely practiced anymore. The task was discarded and replaced with one quite similar; i.e., "Construct and erect forms for concrete walls." Subject matter experts at the Engineer School were consulted and agreed to this change. The remaining discussion will track the task displayed in Figure 2 (Place and Finish Concrete) through the key steps in the simulation test development process. Table 2 shows the task element analysis in terms of its perceptual, cognitive and motor components, and displays the related stimulus variables.

TABLE 2. ELEMENT ANALYSIS TO IDENTIFY COMPONENTS AND STIMULUS VARIABLES

TASK: PLACE AND FINISH CONCRETE (SLAB)

	Critical Elements	Critical Components	Stimulus Variables
1.	Place plastic concrete into forms for slab	Action	
2.	Screed concrete to finish grade	Perception/Action	a. Uniform level. b. Absence of gross low or high spots.
3.	Assist finishing concrete using wood float	Perception/Action	a. Firmness of concrete when pressure is applied. How far foot sinks in. b. Amount of hydration which has taken place. c. Presence or absence of low or high spots as indicated by water pockets. d. Presence or absence of aggregate on the surface.
4.	Assist finishing concrete using steel trowel	Perception/Action	a. Firmness of concrete. b. Presence or absence of water sheen on surface of concrete. c. Sound of trowel when it strikes surface of concrete.

The stimulus variables of critical elements judged to have perceptual components were then analyzed in terms of their importance to task performance. This process is depicted in Table 3. Table 4 depicts the preliminary test mode selection[3/] and stimulus/response similarity.

Finally, Table 5 depicts preliminary judgments as to possible responses on the TV test.

The test items were developed in coordination with personnel at Forts Belvoir and Leonard Wood, based on the element analysis data generated in the simulation procedures. The items were then formatted to a television script (See Appendix C) and reviewed by subject matter experts. Items from Unit 3 (Placing and Finishing Concrete) of the test are shown in Table 6 to illustrate the transference of critical task element and stimulus variables to test items. Certain other items; i.e., 17 and 18 of Unit 3 (Answer Sheet, Appendix C), were added at the request of Fort Leonard Wood personnel to enhance the task continuity and job context realism of the scoreable unit and the test as a whole.

SCOPE AND LIMITATIONS OF THE SIMULATION PROCEDURES

Experience gained during the use of the procedures and the performance of the research has enabled the identification of a number of factors which may limit their effective use.

Availability of Detailed Task Analysis

The procedure assumes as its input, task data in which component skills and knowledges have been identified. The procedure cannot be used effectively if only general task statements are available.

3/ As mentioned earlier, it was contractually determined that videotape (television) would be the presentation mode for the prototype test. Television seemed, to the research team and to ARI, to be the most appropriate medium for simulating the perceptual content of the tasks within MOS 51B. This decision was based on a review and preliminary analysis of the original task list and of the nature of the overall research effort.

TABLE 3. ELEMENT ANALYSIS TO DETERMINE THE IMPORTANCE OF STIMULUS VARIABLES ASSOCIATED WITH A CRITICAL ELEMENT

TASK: PLACE AND FINISH CONCRETE

	Stimulus Variables	Unique	Essential	Importance
Critical Element: Screed Concrete to Finish Grade				
a.	Uniform level of concrete (level with form)	Yes	Yes	Very
b.	Absence or presence of gross high or low spots	Yes	Yes	Very
Critical Element: Assist Finishing Concrete Using Wood Float.				
a.	Firmness of concrete when pressure is applied (How far foot sinks in)	No	Yes	Moderately
b.	Presence or absence of low/high spots as indicated by water pockets	Yes	Yes	Very
c.	Presence or absence of aggregate on surface	Yes	Yes	Very
Critical Element: Assist in Finishing Concrete Using Steel Trowel.				
a.	Firmness of concrete	No	Yes	Moderate
b.	Presence or absence of water sheen on concrete	Yes	No	Moderate
c.	Sound of trowel when it strikes surface	No	No	Moderate

TABLE 4. ELEMENT ANALYSIS FOR PRELIMINARY DETERMINATION OF TEST MODE

TASK: PLACE AND FINISH CONCRETE

	Stimulus Variables	Importance	Realism	Response Similarity	Recommended Format
Critical Element: Screed Concrete to Finish Grade.					
a.	Uniform level of concrete (level with form)	Very	TV/Still	Adequate	TV
b.	Absence or presence of gross high or low spots	Very	TV	Adequate	TV
Critical Element: Assist Finishing Concrete Using Wood Float.					
a.	Firmness of concrete when pressure is applied	Moderate	TV/Still	Adequate	TV
b.	Presence or absence low/ high spots as indicated by water pockets	Very	TV/Still	Adequate	TV
c.	Presence or absence aggregate on surface	Very	TV/Still	Adequate	TV
Critical Element: Assist Finishing Concrete Using a Steel Trowel.					
a.	Sound of trowel striking surface of concrete	Moderate	TV/Slide & Sound	Adequate	TV
b.	Presence/absence water sheen on concrete	Moderate	TV/Still	Adequate	TV
c.	Firmness of concrete	Moderate	TV	Adequate	TV

TABLE 5. ELEMENT ANALYSIS TO DETERMINE POSSIBLE TEST RESPONSE

TASK: PLACE AND FINISH CONCRETE

Stimulus Variable	Job Response	Possible Responses on TV Test
Screed to Finish Grade.		
1. Concrete level with form		1. Show comparator slabs or simply show one slab at a time and examinee makes go/no-go decision on each slab.
a. Above form	a. Remove excess	
b. Below form	b. Add concrete	
2. Absence of gross high or low spots		2. Show comparator slabs or simply show one slab at a time and examinee makes go/no-go decision on each slab.
a. Absence	a. None	
b. Presence high or low	b. Add or remove concrete	
Assist Finishing Concrete Using Wood Float.		
1. Firmness of concrete when pressure is applied		1. Examples of concrete in varying stages of setting.
a. If ok	a. Begin screeding	
b. If too wet	b. Allow more time for concrete to set	
2. Presence/absence of high or low spots as indicated by water pockets	a. If present, float to remove add or remove concrete if necessary	
3. Presence or absence of aggregate on surface	a. If present, float to remove	Visual cue go/no-go
	b. No action if absent	
Assist Finishing Concrete Using Steel Trowel.		
a. Firmness of concrete (pressure)	a. If firm, begin; if not, wait	Show hand on concrete to judge firmness.
b. Presence/absence of water sheen on surface		
c. Sound of trowel striking surface	c. If ringing sound, begin; if dull scrape, wait	Show demonstrator.

TABLE 6. RELATION OF SIMULATION TEST ITEMS TO CRITICAL ELEMENTS AND STIMULUS VARIABLES

TASK: PLACE AND FINISH CONCRETE

Performance Step/ Critical Element		Stimulus Variable	Simulated Test Item
Assist finishing concrete using wood float	a.	Firmness of concrete when pressure is applied	Item 25. Is this concrete ready to be floated?
	b.	Amount of hydration which has taken place	(VISUAL: PAN OF SURFACE, showing adequate hydration. HOLD on demonstrator putting his foot into surface to check firmness.) A. Yes B. No C. The information is insufficient.
	c.	Presence or absence of low or high spots as indicated by water pockets, ridges	Item 27. Which slab has been properly floated? (VISUAL: PAN OF SURFACE, of four comparator slabs, three show water spots, ridges or aggregate. One is proper.)
	d.	Presence or absence of aggregate on surface	A. B. C. D.
Assist finishing concrete using steel trowel	a.	Firmness of concrete	Item 28. Look at this concrete. Is it ready for <u>first</u> troweling?
	b.	Presence or absence of water sheen on surface	(VISUAL: PAN OF SURFACE, to show slight water sheen. HOLD on demonstrator as he applies trowel.) AUDIO: Sound of trowel ringing on surface.
	c.	Ringing sound of trowel when it strikes surface	A. Yes B. No C. The information is insufficient.

Availability of Appropriate Perceptual Content

It is obvious that most human behavior occurs within the context of the perceived environment; hence most tasks are said to have a perceptual component. The simulation procedure, however, is designed to identify situations in which the perceptual component is the critical component of appropriate task oriented behavior. Such situations are in fact relatively rare. Typically, people learn to recognize job related stimuli and errors occur as a consequence of inaccurate decisions or motor responses. Baldwin (1971) discusses industrial training and notes the importance of perceptual learning. For example, the TV repairman must learn to identify hums at 60 Hz, 120 Hz, and 15,750 Hz, and he must learn to identify odors such as burned insulation or a burned out transformer. Baldwin points out, however, that these stimuli are sufficiently discrete that testing a graduate is unnecessary.

An example of an area in which perceptual judgment continues to be a major factor in performance after training is the auto mechanic's task of using feeler gauges to adjust valves or points. Here proprioceptive sensitivity remains the key element to accurate performance.

In the present application of the procedures, it was difficult to select appropriate test content because tasks associated with levels one and two of the 51B MOS have few critical perceptual requirements. For example, the task of jointing and sharpening a saw is described (in the technical manual) as having considerable perceptual content, but in practice, carpenters rely on learned motor responses to assure that they have properly performed the task. Also, the determination of a level surface would be a perceptual problem if no aids were available, but it becomes a cognitive problem when the task is to decide whether the level is sufficient when the carpenter recognizes that the edge of the bubble in the carpenters level is slightly over the edge of the line.

Experience of the User of the Procedures

The algorithm was intended for use by a relatively unsophisticated test developer. Experience now suggests that some concepts and procedures involved are relatively strange to this target individual and that more experience or training may be a prerequisite for proper use.

Availability of Manpower to Support User

The procedure involves the frequent use of from 2 to 5 subject matter experts, as well as the requirement for a media expert and for examinees during the assessment of test validity and reliability. The research team experienced difficulty in obtaining part of the support specified, and this is not attributed to a lack of cooperation with the team in particular. It is noted that few SQTs are fully validated in complete accordance with present doctrine. This is probably the most limiting factor with respect to the use of the procedures in the Army test development environment; in fact, the lack of manpower support was the principal reason that the procedures were not completely followed in this research.

PRODUCTION AND EVALUATION OF THE SIMULATION TEST

Production of the simulation test took place at Fort Leonard Wood, MO, where both MOS 51A and 51B are taught. Instructors from the courses were used as "actors" in the test since they were already experts in the tasks and it was felt they could more easily adjust their work to suit the special requirements of recording the effort on videotape.

PRODUCTION OF THE TEST

The simulation test consisted of three parts, each representing a scoreable unit. These units were:

- Unit 1: Hand Tool Maintenance and Material Preparation (11 items)
- Unit 2: Erecting Wall Forms (9 items)
- Unit 3: Placing and Finishing Concrete (7 items)

The first and third units consisted of four alternative multiple-choice items; the second, two alternatives ("correct" and "incorrect"). In addition, the second unit contained a number of unalerted safety violations, and examinees were asked to identify the type of safety violations by a code number as follows:

1. Failure to ground electric tools or equipment properly.
2. Failure to wear protective gear when necessary.
3. Use of tool in a hazardous manner.
4. Unsafe vehicle operating procedures.

The logistics of producing the test made a meaningful review cycle impossible prior to validation of the test. For example, the training schedule at Fort Leonard Wood was such that the intended sample population would be available for only one day and on a date which could not be adjusted without a seven-week delay. Therefore, it was necessary to carry out the test validation on the Tuesday following the Friday completion of the draft test.

However, the day before the test was administered, it was shown to a group of seven instructors from the carpentry course, as part of the validation process. There was agreement with the scoring key (correct and incorrect classification of alternatives).

IDENTIFICATION OF EXTERNAL CRITERIA

As noted in the background section of this report, SQT items were not available as external criteria for purposes of validation. Thus it was not possible to validate performance on each critical task element in the simulation test against performance of the identical critical task element in a performance test. The only available performance related data for purpose of validation were ratings which were associated with each man's performance in Carpentry and Utility Worker Training. The module scores: (1) were based on more general observations than would be required for assessment of performance at the level of the critical task element, (2) included observations of task elements which were perhaps related to, but not part of the simulation test, and (3) did not include the observation of all task elements which were included in the simulation test. The soldiers included in this sample had been observed in the performance of the course modules from two to seven weeks prior to their performance of the simulation test; all were trainees who had received further training during this period.

For each unit of the test, the one or more module ratings that most closely matched the content of the test was used. Thus for Unit 1 of the simulation test, which consisted largely of carpentry related tasks, the "Building Construction" module ratings were used for Carpenter trainees and the "Carpentry" module ratings were used for utility worker trainees. Both modules contained work sample performance tests or identification problems on:

(1) Carpenter tool maintenance; i.e., the sharpening and jointing of hand saws.

TABLE 7. RELATION OF EXTERNAL CRITERIA TO TEST ITEMS

Criteria	Simulation Test Item(s)
Handsaw sharpening and jointing	1, 2, 3, and 9
Construction layout	4, 5, 6, 7, and 8
Material identification	10

(2) Material identification; i.e., proper lumber, nails, etc., for the job.

(3) Construction layout; i.e., rudimentary problems in determining proper lengths of lumber and proper angles.

Table 7 shows the criteria/test item correspondence. However, the training brigade maintained only the overall module ratings, so that performance ratings on item-specific criteria could not be obtained.

For Unit 2, the "Building Construction" module ratings were used and the material identification and construction layout portions appeared to offer adequate correspondence.

A combination of "Carpentry" and "Masonry" module ratings were used for utility worker trainees.

Because the carpenter group had no "Masonry" module, Unit 3 (Placing and Finishing Concrete) was administered only to the utility workers group. The ratings on the "Masonry" module were used as criteria. The module test included the actual placing and finishing of a small concrete slab as well as oral questions on when and how to vibrate concrete.

The data obtained consisted of numerical ratings for each trainee in each course module, as well as the number of repetitions of the module. In identifying criterion ratings, the last rating obtained by the trainee was used. For example, one trainee took a module three times and obtained ratings of 47, 50, and 60. Sixty was used as the criterion score. It was reasoned that this was a truer index of his performance than either an earlier rating or an average of his ratings.

TABLE 8. DESCRIPTION OF EXPERIMENTAL GROUPS AND CRITERIA*

Test Unit	Group	n	Criteria
Unit 1: Hand Tool Characteristics and Material Preparation	Carpenters	24	Building Construction Module
	Utility Workers	23	Carpentry Module
Unit 2: Erecting Wall Forms	Carpenters	24	Building Construction Module
	Utility Workers	23	½ (Carpentry & Masonry Modules)
Unit 3: Placing and Finishing Concrete	Utility Workers	23	Masonry Module

*Applies to both simulation (television) and written tests.

The experimental design also called for correlating the simulation test with a written SQT component. In the absence of the written component, the research team constructed a written test which matched the simulation test on a task-for-task level, but not a critical element-for-critical element level.

SELECTION OF MATCHED GROUPS

Because of the small number of subjects available for this study, (See Table 8) "matched" groups of carpenter and utility worker trainees were used rather than random samples from the available population. One group received the simulation test followed by the written test, while the other received the written test followed by the audio-visual test.

The matching variable selected for the carpenter trainees was their rating on the Building Construction module. The matching variable selected for the utility worker trainees was their rating on the Carpentry module.

Each group (i.e., carpenters and utility workers) was ranked with regard to the matching variable and then assigned to groups A and B, as follows: AB, BA, AB, BA, etc. Thus for every pair, the higher was alternately assigned to groups A and B.

TEST ADMINISTRATION PROCEDURES

The tests were administered to the carpenter trainees in the morning and the utility worker trainees in the afternoon.

Each group reported to the classrooms that were used for simulation test administration. Students were assigned to one of the two classrooms. One of the two test administrators stayed with each group.

The group taking the simulation test was told that it was an experimental test, and that their scores would not be given to their units. They were asked to do their best, and to guess at the answers they did not know.

Answer Sheets (Appendix C) and a diagram of the wall form (Figure 1, Appendix D), were distributed.

Similar instructions were given to the group taking the written test (Appendix D). A separate answer sheet was not required, since examinees circled their answers in the test booklet. Since the simulation test took longer to administer, students completing the written test were allowed to walk outside of the building, while waiting to exchange places with the simulation test group.

As soon as the simulation test was over, the groups exchanged places and each was given the alternate test.

POOLING OF TEST DATA

Each test was administered to both groups (i.e., carpenter trainees and utility worker trainees). In addition, each group was divided into two subgroups with regard to testing sequence (i.e., one received the simulation test followed by the written test, and the other which received the tests in reverse order).

To obtain a larger n for the validation process, it was desirable to pool the test results obtained for the two groups of trainees and the two orders of presentation. The results of a two-factor analysis of variance, using total test scores, is shown in Tables 9 and 10. It can be seen that order of presentation (sequence effects), group membership, as well as the interaction between the two factors did not result in significant variance. Accordingly, data from both groups as well as from both orders of presentation was pooled.

TABLE 9. ANALYSIS OF VARIANCE FOR SIMULATION (TELEVISION) TEST SCORES

Source	df	MS	F	P
Order of Presentation	1	26.01	2.29	Ns
Carpenters versus Utility Workers	1	27.13	2.39	Ns
Interaction	1	14.36	1.26	Ns
Within Groups	43	11.37	--	--

TABLE 10. ANALYSIS OF VARIANCE FOR WRITTEN TEST SCORES

Source	df	MS	F	P
Order of Presentation	1	5.73	1.01	Ns
Carpenters versus Utility Workers	1	22.18	3.90	Ns
Interaction	1	7.55	1.33	Ns
Within Groups	43	5.68	--	--

Another source of evidence for the comparability of the two groups was found by an inspection of the AFQT scores of the trainees. The average AFQT score for the carpenter trainees was 57.45 with an S.D. of 20.86. For the utility worker trainees, the average AFQT score was 48.90 with an S.D. of 24.49. The difference between these group averages was 8.55 and a test for the mean differences yielded $t = 1.26$, which was not significant.

TABLE 11. MEANS AND STANDARD DEVIATIONS OF RATING SCORES ON CRITERION MODULES BY ASSOCIATED SIMULATION TEST UNIT AND GROUP

Simulation Test Unit	Group	n	Criterion Module	Ratings	
1	Carpenters (A)*	10	Building Construction	87.80	8.34
	Utility Workers (A)	12	Carpentry	77.50	16.03
	Carpenters (B)*	14	Building Construction	84.36	10.45
	Utility Workers (B)	11	Carpentry	82.73	12.32
2	Carpenters (A)	10	Building Construction	87.80	8.34
	Utility Workers (A)	12	(Carpentry + Masonry)	81.83	15.63
	Carpenters (B)	14	Building Construction	84.36	10.45
	Utility Workers (B)	11	(Carpentry + Masonry)	86.55	10.88
3	Utility Workers (A)	12	Masonry	85.83	16.76
	Utility Workers (B)	11	Masonry	90.00	14.14

NOTE: Group A were given the simulation test first
Group B were given the written test first

PERFORMANCE ON COURSE MODULES

Ratings on course modules were used as the performance criterion in validating the simulation test. Table 11 presents the means and standard deviations of the distributions of ratings for the samples of carpenters and utility workers relative to: (1) the simulation test unit associated with the specified modules, and (2) the testing sequence for the two groups of examinees. The data in Table 11 are interpreted as follows: (1) mean performance ratings on each of the three criterion modules (i.e., building construction, carpentry, and masonry) were quite high, and (2) the examinees were still somewhat heterogeneous in their performance levels as evidenced by the amount of variance associated with the means.

TABLE 12. MEANS AND STANDARD DEVIATIONS OF PERCENTAGES OF CORRECT RESPONSES ON THE WRITTEN TEST FOR CARPENTERS AND UTILITY WORKERS

	Carpenters		Utility Workers		Pooled	
Test Unit	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	53	16	43	19	49	18
2	50	21	42	18	46	20
3	47	19	49	15	48	12
Total Score	51	12	45	11	48	12

PERFORMANCE DATA ON WRITTEN TESTS

The written test was used as a basis for assessing the relative value of written or simulation tests. Examinee performance on the written test is summarized in Table 12. The means of the percentages of correct responses on each of the three units are quite low, indicating that the test was fairly difficult. The magnitude of the standard deviations suggests a considerable range in examinee knowledge levels.

PERFORMANCE DATA ON THE SIMULATION TEST

The means and standard deviation of the distributions of test scores (percentages of correct answers) offered by the carpenter and utility worker groups are given in Table 13. These means are close to the 50 percent, which from a psychometric point of view is good since it offers the greatest possibility of maximum item discrimination (Ebel, 1972). From the Army's perspective, these tests are somewhat too difficult. The magnitude of the standard deviations suggests a considerable range in examinee abilities.

UNIT AND OVERALL TEST VALIDITY

The estimation of test validity for both the simulation test and the written test was accomplished by correlating unit and overall test scores with selected module ratings. The modules that were selected for validating each unit and the associated rationale have been identified earlier in this report.

TABLE 13. MEANS AND STANDARD DEVIATIONS OF PERCENTAGES OF CORRECT RESPONSES ON THE SIMULATION TEST FOR CARPENTERS AND UTILITY WORKERS

	Carpenter (n = 24)		Utility Workers (n = 23)		Pooled (n = 47)	
Test Unit	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	54	23	59	17	56	20
2	44	19	51	16	48	18
3	58	21	63	13	61	18
Total Score	52	15	57	10	54	13

The correlations are presented in Table 14. Both the simulation test and the written test provided significant correlations with the respective criteria for Unit 1 and for the test as a whole. The contribution of the variance of Units 2 and 3 for the respective tests is minimal, so it must be concluded that only Unit 1 is providing variance to the test as a whole.

These results seem to represent some improvement over Cockrell's (1976) results. His overall correlation of the written test was .329 ($p < .01$), but his overall correlation of the simulation test with his criterion was only .235 (NS).

However, it must still be acknowledged that the magnitude of the correlations indicate that both tests were of moderately low validity.

ASSESSMENT OF ATTITUDES TOWARD SIMULATION

Examinee attitudes toward simulation testing were assessed via a questionnaire.

In developing the questionnaire, the research team adapted items found in the Procedures for Validating Skill Qualification Tests (Hirshfeld, Young and Maier, 1975), informal questionnaires obtained from the Army service schools, as well as questionnaires found in academic literature. The questionnaire contained both structured (scaled) items and open-end items. It was

TABLE 14. CORRELATION OF SIMULATION AND WRITTEN TEST SCORES WITH RATINGS ON CRITERION MODULES

Test and Unit	Pearson	r	n	p
Simulation Test				
Unit 1		.427	47	.01
Unit 2		.145	47	NS
Unit 3		.068*	47	NS
Total Score		.350	47	.02
Written Test				
Unit 1		.448	47	.01
Unit 2		.021	47	NS
Unit 3		.182*	47	NS
Total Score		.430	47	.01

*Correlation based upon utility worker trainees only, since criterion (Masonry Module Scores) was not available for carpentry trainees.

designed to assess and contrast attitudes toward both the simulation and written tests.

STRUCTURED ITEMS

These items were constructed to provide a scaled response on a four-point scale. The data were recorded separately for examinees who were tested in the morning versus those tested in the afternoon because of the differences reported by Cockrell (1976).

The data presented in Table 15 reveals that overall, the examinees felt that the A/V simulation test was slightly fairer, slightly more interesting, and very slightly less difficult than the written test. The data further reveal that on seven of eight items (Items 1, 2, 4, 5, 6, 7, 9, 10) in which one end of the scale represented a positive feeling, the average response (mean) fell between mid-scale and the positive end. The one item in which the mean fell

TABLE 15. SCALED QUESTIONNAIRE DATA

1. To what extent was either test a fair measure of your ability to perform in your MOS?

Television Test		Written Test	
1 Extremely fair		1 Extremely fair	
2 Very fair	$\overline{X}$ = 2.36	2 Very fair	$\overline{X}$ = 2.64
3 Somewhat fair		3 Somewhat fair	
4 Not fair		4 Not fair	

2. How interesting was either test?

Television Test		Written Test	
1 Extremely interesting		1 Extremely interesting	
2 Very interesting		2 Very interesting	
3 Somewhat interesting		3 Somewhat interesting	
4 Not interesting	$\overline{X}$ = 2.44	4 Not interesting	$\overline{X}$ = 3.11

3. How difficult was either test?

Television Test		Written Test	
1 Extremely difficult		1 Extremely difficult	
2 Very difficult	$\overline{X}$ = 3.27	2 Very difficult	$\overline{X}$ = 3.24
3 Somewhat difficult		3 Somewhat difficult	
4 Not difficult		4 Not difficult	

4. Overall, to what extent were the visuals (pictures, graphics, titles, etc.) clear in the television test?

1 Extremely clear	
2 Very clear	$\overline{X}$ = 2.28
3 Somewhat clear	
4 Not clear	

5. In the case of television test, to what extent was the narration easy to understand?

1 Extremely easy to understand	
2 Very easy to understand	$\overline{X}$ = 1.71
3 Somewhat easy to understand	
4 Not easy to understand	

TABLE 15. SCALED QUESTIONNAIRE DATA (Continued)

6. In the case of the television test, to what extent was the answer easy to use?

1 Extremely easy to use
2 Very easy to use
3 Somewhat easy to use
4 Not easy to use

$\overline{X}$ = 2.08

7. Overall, did you have enough time to answer the questions to the television test?

1 More than enough time
2 Enough time
3 Barely enough time
4 Not enough time

$\overline{X}$ = 2.34

8. What is your feeling about the overall pace (rate of presentation) of the television test?

1 The pace was much too slow
2 The pace somewhat too slow
3 The pace was somewhat too fast
4 The pace was much too fast

$\overline{X}$ = 2.49

9. What is your feeling about the overall selection of items (situations) for the television test?

1 The items were extremely well chosen
2 The items were very well chosen
3 The items were fairly well chosen
4 The items were poorly chosen

$\overline{X}$ = 2.71

10. From where you are sitting, how well were you able to see the television screen?

1 Extremely well
2 Very well
3 Fairly well
4 Not well

$\overline{X}$ = 1.60

on the negative side of the midpoint (item 9) pertained to the selection of test items. Item 8 referred to the pace, and the responses were close to the midpoint, indicating a pace that was neither too fast nor too slow. The tenth item (item 3) referred to test difficulty, and responses indicated that the test was judged to be somewhat difficult.

OPEN-END ITEMS

The constructed items yielded only one consistent finding, namely that Unit 2 was a little confusing. Some representative comments to the various open-end questions are given below:

(1) "Can you recall any specific items (situations) in the television test that are confusing? If so, describe the item(s) in a few words".
   (a) I think that in Unit #2 they went too fast. You'd write on answer sheet, look up, and the next question was almost over.
   (b) On Unit #2, the skipping around was a little confusing. Also looking for safety hazards was hectic.
   (c) The building of the wall form - was extremely hard to follow the procedures.

(2) "Do you have any additional comments on the television test"?
   (a) I thought it was a good test. I think those kind of tests would be alright!
   (b) There wasn't enough verbal instruction during the test.
   (c) Many of the pictures (alternatives--Ed.) looked the same.

(3) "Can you recall any specific items in the written test that were confusing? If so, describe the item(s) in a few words".
   (a) The parts about filing a saw.
   (b) The diagram of the wall form.

(4) "Do you have any additional comments on the written test"?
   (a) The written test to me wasn't clear enough.
   (b) It was a lot easier to understand than the television and you could work at your own speed.
   (c) I think it is better to show the TV because of the people that can't read.

COMPARATIVE COSTS OF VIDEOTAPE SIMULATION TESTS

It was intended that the actual costs of developing, validating, and administering the prototype test would be compared with the actual costs of developing, validating, and administering the same test in a hands-on performance and written format. As previously stated, no hands-on or written tests existed and the Army was unable to supply firm data as to development and administration costs of any SQT component. Nor was cost data available from the only other two known television simulation tests. Therefore, certain costs pertaining to the development and administration of hands-on and written versions of the prototype simulation test were estimated by the research team's test development staff who recently completed an SQT for the Ordnance School at Aberdeen Proving Grounds, MD. In addition, because the simulation test was a prototype and not intended for actual use, the administration costs for it were estimated. The assumptions under which these estimates were made are stated as follows:

Assumptions

(1) One videotaped test is given to a maximum of 15 soldiers in one administration; the test is 60 minutes long and requires 1-1/2 manhours to administer.

(2) One hands-on test is given to one soldier in one administration; the test may take up to two hours and requires one rater to administer.

(3) One written test is given to a maximum of 50 soldiers in one administration. The test is 40 minutes long and requires 1 manhour to administer. Material cost (paper) for this administration is $25.00.

(4) Materials used in the hands-on test, i.e., lumber, concrete, etc., are considered as consumables in that they would not be re-used for testing purposes.

(5) For the purposes of this analysis, each test site would receive two video cassette copies (one for back-up) of the test. One cassette may be played 50 times before its quality is degraded. (This is a conservative estimate.) Therefore, one cassette could conceivably test 750 soldiers (50 replays times 15 soldiers).

(6) Costs for administering all three tests are linear; i.e., the costs of administering the same test to the same number of soldiers will be the same for the first test and the 1,000th test, provided, of course, assumptions 1 through 4 are met.

(7) Because dollar amounts seemed the most appropriate common denominator, it was necessary to ascribe an hourly rate to a "typical" NCO scorer/test administrator. For purposes of this analysis, an E-6 (Staff Sergeant) was chosen, with six years service, no dependents and no proficiency pay. His hourly rate was estimated as follows:

BASE PAY	$655.00/mo.
ALLOWANCE FOR QUARTERS	$117.00/mo.
ALLOWANCE FOR SUBSISTENCE	$ 79.50/mo.
CLOTHING ALLOWANCE	$ 7.00/mo.
TOTAL	$859.00/mo.
PER HOUR	$ 3.30/hr.

Military pay is based on a 30-day month and there is no stated standard work day. With this in mind, the typical NCO was arbitrarily assigned a 10-hour work day with allowance for a minimum of four non-work days per 30-day period. This reduces to 260 productive hours per pay period, at $3.30/hour.

Table 16 shows actual expenditures of professional manhours for development and validation of the videotaped test and estimated expenditures for the hands-on and written tests.

Costs involved in task analysis, task selection, coordination, media selection, and the like are not included as it is assumed these costs would be roughly the same regardless of test format. Travel costs are not considered as they are extrinsic to the development process and would probably not occur in an operational (as opposed to a research) mode.

Material costs for the videotaped test are as shown in Table 17.

Equipment costs are considered as sunken costs and therefore not relevant to the analysis. The Army is upgrading and expanding its videotape

TABLE 16. DEVELOPMENT AND VALIDATION EXPENDITURES

	Professional Manhours		
Task	Videotape	Hands-On*	Written*
Development	848	120	16
Validation	80	40	8
Total	928	160	24

*Estimated

production capability and the equipment and operating personnel are presently used in the production of training and command information materials. Therefore, the equipment costs in production of videotaped tests are really opportunity costs; that is, the cost/benefit of producing tests versus training materials.

The alternative uses of existing capital equipment are policy decisions and beyond the scope of this report. It will be demonstrated, however, that the relatively low administration costs of videotaped tests make it a feasible test delivery system when compared to the costs of administering similar hands-on tests. The savings in administration outweigh the admittedly high costs of development and production.

Material costs used in validation of the hands-on test would be the same (less the expense of the videotape), or about $200.00. As can be readily seen from the preceding data, development costs of the videotaped tests are significantly greater than those of the hands-on version (848 professional manhours compared to 120 professional manhours). This supports Cockrell's observation (page 13). However, administration costs demonstrate the feasibility of the videotaped version.

Costs per soldier incurred in administering the hands-on test (assumption 2) are computed as follows:

TABLE 17. VIDEOTAPED TEST MATERIAL COST (DEVELOPMENT)

Concrete	$ 25.00
Lumber	$175.00
Videotape Cassettes	$500.00
	$700.00

PER SOLDIER COSTS OF HANDS-ON TEST ADMINISTRATION =

$$\frac{\text{(MATERIAL COSTS) + (NCO HOURS X \$3.30)}}{\text{NUMBER OF EXAMINEES TESTED IN ONE ADMINISTRATION}}$$

Based on the estimates, the hands-on test would cost $\frac{(\$200) + (2 \times \$3.30)}{1}$ or $206.60 per soldier tested.

The same formula is used for computing the per soldier costs of administering the written test (assumption 3). Thus, the test cost $\frac{(\$25) + (1 \times \$3.30)}{50}$ or slightly less than $0.57 per soldier tested.

On the other hand, the cost per soldier of administering the videotaped version (assumption 1) can be stated as the cost of the tapes divided by the total number of examinees (up to 750, assumption 5), plus the total NCO hours expressed in dollars, also divided by the total number of examinees. The NCO dollar figure is computed by dividing the total number of examinees by 15 (assumption 1), counting the remainder as the next whole unit, and multiplying that figure by $3.30 (assumption 7).

The formula can be stated as follows:

$$\text{PER SOLDIER COST OF VIDEOTAPED TEST ADMINISTRATION} = \frac{\text{COST PER TAPE X NUMBER OF TAPES}}{\text{NUMBER OF EXAMINEES}} + \left[\left(\frac{\text{NUMBER OF EXAMINEES}}{15}\right)^{*} \times \$3.30\right]$$

*Rounded up to whole number.

Based on the estimates, then, the per soldier cost of the prototype test administration would be (number of examinees = 47):

$$\left(\frac{\$30 \times 2}{47}\right) + \left(\frac{47}{15} \times \$3.30,\right) \text{ or } \$1.28 + \$13.20, \text{ or } \$14.48 \text{ per soldier tested.}$$

The comparative test administration costs per soldier are:

Hands-on test	$206.60
Videotaped test	$ 14.48
Written test	$ 0.57

As mentioned earlier, the videotape production equipment costs have not been considered here. However, it was felt that an equipment configuration which is considered as the minimum necessary to produce the prototype should be described. The equipment detailed in Table 18 could be purchased for approximately $60,000.00.

The preceding discussion deals, of course, only with the cost advantages of the videotaped simulation test. In this light, it supports Cockrell's contention, "The tests can be produced at a reasonable cost ..." (page 16) and that, "television testing is far less cumbersome and costly than hands-on testing" (page 13). It is recognized that decisions based purely on monetary cost savings ignore the many other aspects that contribute to the value of a given alternative. If the simulation test does not measure knowledge and skills with acceptable accuracy, the price may be right, but its value is questionable.

TABLE 18. VIDEOTAPED PRODUCTION EQUIPMENT CONFIGURATION

PORTABLE EQUIPMENT	QUANTITY
Portable vidicon color cameras with tripods	2
Portable video cassette color recorder with microphone	1
Portable lighting kit	2
STUDIO EQUIPMENT	
Video cassette recorder/players with edit capability	2
Edit controller	1
Special effects console	1
Time base corrector	1
Monitor (color)	2
Audio mixer	1
Audio tape recorder	1

LESSONS LEARNED

The following is a general discussion of judgments and observations made by the research team during the course of the simulation test construction process.

LIMITATIONS AFFECTING THE VALIDATION OF THE SIMULATION TEST

First, test items within units were not adequately revised following review of the completed prototype test by subject matter experts at Fort Leonard Wood. This was due to scheduling problems discussed earlier. On some items, there was disagreement among the experts as to what the prescribed doctrine was, even when the doctrine was spelled out in the relevant technical manuals. This lack of agreement among technical experts seems to

extend to the construction of the standard SQTs as well. It is exacerbated somewhat when television is the testing mode because making changes to videotape is more cumbersome than making changes to written tests or hands-on test scoring procedures. The "lesson" seems to be, to insist on a more intensive and higher level review of the test in its storyboard and scripting stages.

A second factor pertains only to Unit 2 of the simulation test. This unit used a relatively unalerted response format, in which the examinee had to record an event which might occur anytime during a unit of up to two minutes. The examinees were not familiar with this response format, and they reported some confusion in their questionnaire responses. The testing session did not include any example responses for the Unit 2 response format. Such warm-up may have alleviated the problem. It is noted that Cockrell (1976) indicated the need for examinee practice in responding to televised test items.

Third, and perhaps most importantly, the criterion measure was only indirectly related to the simulation test item. This factor alone should immediately be expected to reduce the correlation to a moderate level at most. The experimental design called for a validation on an item-by-item basis. Since appropriate performance measures were not supplied by the Army, the simulation test was validated against a measure of performance from a large training module. The module may or may not have included the specific skills which were tested in the simulation test, and the observations may have been made up to a month prior to the administration of the simulation test.

PROBLEMS INHERENT IN PRESENT APPLICATION OF A/V SIMULATION TESTS

The overall purpose of this research effort was to validate a promising method of audio-visual simulation to test perceptual job content in a real world application. Previous discussion has identified a few problems encountered, most of which were not inherent in the situation. At this time, it is appropriate to specify problems which would have occurred even if logistic problems such as scheduling and availability of subject matter experts had not occurred. These problems are addressed because they should be considered in other attempts to specify appropriate applications of A/V simulation testing.

PERCEPTUAL CONTENT IN JOB

The present application, by design, was limited to perceptual or perceptual related psychomotor task components. The first obvious requirement for this application was that there be appropriate perceptual or perceptual related tasks to test. In fact, few tasks in the levels one and two of MOS 51B were appropriate. While all three units of the simulation test addressed perceptual components of tasks, it is recognized that for Unit 1 in particular, cognition was more critical to accurate performance than perception. (Further errors in actual task performance are most often psychomotor). For example, an incorrect response to an item showing four positions for pointing a hand saw was more likely related to not "knowing" the right position than to not "discriminating" the alternative positions. In MOS 51B, critical perceptual tasks occur more at the supervisory level where there are inspections of work in process or accomplished.

SIMILARITY OF TEST AND JOB RESPONSE

In a preceding discussion of test validity, the importance of response variables was noted. In the present application, levels 1 and 2 tasks typically involve psychomotor responses and the responses are not quite as alerted as in the multiple-choice format. (The response of the supervisor which is often a verbal response, would not be as discrepant.) In the present application, disparity between the job response and available test responses presented a problem.

GENERAL CONSIDERATIONS CONCERNING THE USE OF TELEVISION SIMULATION

MOTIVATIONAL/ATTITUDINAL

Videotape seems to have an advantage over performance and paper-and-pencil tests insofar as its acceptability to the examinee is concerned. This is an important criterion in the consideration of test formats. Bloom (1970) concludes that student measurement can have both positive and negative effects and that the person being evaluated will always respond to evaluation

in terms of the perceived fairness. This perceived fairness is enhanced through television because the test developer is able to take advantage of the "transfer effect" as potential examinees are already highly receptive to the medium. Thus, the television test builds on habit patterns already firmly established in the examinee. Cockrell (1976) argues for continued interest in the use of television as the stimulus input in synthetic performance testing. He lists three reasons given by examinees for preferring television testing:

> "1. Scoring is fairer and not dependent upon the whims of the test administrator.
> 2. Testing is faster and not so drawn out.
> 3. In television testing, no one is shouting at you and ordering you around."

The questionnaire results reported both favorable attitudes towards the TV test and more favorable attitudes toward it than toward a written test. Thus, although the present TV application was justified only in terms of perceptual test content, developers of SQTs might wish to consider these other factors.

RESOLUTION/ACUITY

When judgments are based on fine perceptual discriminations, such as the presence or absence of a light water sheen on concrete or minor differences in the color or grain of types of lumber, television may not be able to faithfully reproduce these visual cues. There is no standard picture quality from one television set to the next; hence, there is no assurance that visual stimuli simulated faithfully on a studio monitor will not be obliterated by a set with poorer resolution in the field.

Because TV or any A/V medium which may be selected for simulation is limited with respect to the stimuli that can be faithfully reproduced, it is necessary to carefully analyze the perceptual content before deciding that simulation is an acceptable testing mode.

COSTS OF A/V SIMULATION VERSUS WRITTEN AND PERFORMANCE TESTING

The costs of administering a television test are significantly less than those of a performance test and only slightly higher than those of a written test. The major costs of the television test are incurred during the production stage. These costs can be lessened through the early involvement of media experts in the test development process, so that exceptionally costly or time consuming segments can be kept to a minimum. Production costs can also be lessened through an intensive review of the test in the scripting stage, so that editing and "reshooting" time is minimized.

SIMULATION OF DYNAMIC TASKS

Television is able to realistically simulate situations where the viewer has to make judgments based on his observations of relatively complex, dynamic events in which many variables are changing. The television portion of the Military Policeman SQT, for example, has a great deal of face validity in its presentation of many such situations as test items. Through the use of television, the test developers were able to simulate, with adequate realism, the scene of a burglary, complete with scattered pieces of evidence which the examinee was asked to identify and note (the same response as required on the job). Such an item could not have been described verbally without extreme overcuing. Likewise, the MPs were able to simulate a situation which required that the examinee locate a particular suspect, based on brief descriptions, from among a moving crowd of people, while the camera moved through a neighborhood in a patrol car. These items called for judgments based almost entirely on fine perceptual discriminations and demonstrated the true appropriateness of the medium.

Soldiers at higher, supervisory skill levels must frequently make judgments based on their observations of dynamic events, such as the crew performance involved in many construction tasks. At the higher skill levels, however, many times the proper response to a given stimulus is the judgment. The supervisor's task may be to observe a process and note errors and inconsistencies in the process. If this process is a dynamic one, where changes in the states of the variables overlay one another and do not necessarily occur in a prescribed sequence, then a dynamic medium such as television is necessary if the process is to be faithfully simulated.

The few truly perceptual judgments made by soldiers at skill levels 1 and 2, however, are generally based on stimuli which are static or at least flow in an easily predictable sequence. This static, or sequential characteristic of task elements at the lower skill levels, makes television less appropriate as a test medium in those cases.

CONCLUSIONS

TASK SELECTION SIMULATION PROCEDURES

1. The application of the procedures enabled the selection of the more appropriate tasks and task components from a specified field of tasks critical to MOS 51A and 51B.

2. Use of the simulation procedures requires a greater expenditure of human resources than may typically be present in a test development agency.

APPLICABILITY OF A/V SIMULATION

1. The fundamental question of the applicability and validity of A/V simulation to test perceptual content was not conclusively answered because of a number of problems discussed in the text of the report.

2. The use of television as a simulation means, strictly for testing the perceptual content of lower skill level motor tasks such as those within the Carpentry and Masonry MOS, appears somewhat limited; there appears, however, to be a decided favorable attitudinal bias, on the part of the test takers, towards television testing.

REFERENCES

Baldwin, T. S., "Evaluation of Learning in Industrial Education," in Bloom, B. S. et al (Ed), Handbook on Formative and Summative Evaluation of Student Learning, McGraw-Hill, Inc., New York, 1971.

Bennett, C. A., "Toward Empirical, Practicable, Comprehensive Task Taxonomy," Human Factors, 1971, 13(3).

Berliner, C., Angell, D., and Sheaver, J. W., Behaviors, Measures and Instruments for Performance Evaluation in Simulated Environments. Paper presented at the Symposium and Workshop on the Quantification of Human Performance, Albuquerque, New Mexico, August 1964.

Bloom, B. S., "Toward a Theory of Testing Which Includes Measurement-Evaluation-Assessment," in Wittrock, M. and Wiley, D. (Eds), The Evaluation of Instruction. Holt, Rinehart and Winston, New York, 1970.

Cockrell, John, Television as Stimulus Input in Synthetic Performance Testing, Experiment I, US Army Research Institute for the Behavioral and Social Sciences, Ft. Knox Field Unit, Kentucky, July 1976.

Cotterman, T. E., Task Classification: An Approach to Partially Ordering Information in Human Learning (WADC TN 58-374), Wright-Patterson Air Development Center, Ohio, 1959.

Ebel, Robert L., Essentials of Educational Measurement. Prentice-Hall, Englewood Cliffs, N. J., 1972.

Fleishman, E. A., "Toward a Taxonomy of Human Performance," American Psychologist, 1975, 30(12).

Foley, J. P., Jr. Evaluating Maintenance Performance: An Analysis, (AFHRL-TR-75-61). Air Force Human Resources Laboratory, Wright-Patterson Air Force Base, Ohio, October, 1974.

Folley, J. D., Jr., Development of an Improved Method of Task Analysis and Beginnings of a Theory of Training (Report 1218-1). US Naval Training Devices Center, Orlando, Florida, June 1964.

Gagne, R. M., "Human Functions in Systems," in Gagne, R. M., (Ed.), Psychological Principles in System Development. Holt, Rinehart and Winston, New York, 1962.

Hirshfeld, Stephen F., Young, Douglas L. and Maier, Milton H., Procedures for Validating Skill Qualification Tests (Draft). US Army Research Institute for the Behavioral and Social Sciences, July 1976.

Laabs, Main, Abrams, and Steinnemann, A Personnel Readiness Training Program: Initial Project Developments (Special Report 75-8). US Navy Personnel Research and Development Center, San Diego, California, April 1975.

Osborn, William C., Campbell, Ray C., Ford, J. Patrick, Hirshfeld, Stephen F., and Maier, Milton H., Handbook for the Developers of Skill Qualification Tests. US Army Research Institute for the Behavioral and Social Sciences, Alexandria, Virginia, 1977.

Osborn, William C., An Approach to the Development of Synthetic Performance Tests for Use in Training Evaluation, (HumRRO Professional Paper 30-70). Human Resources Research Organization, Alexandria, Virginia, December 1970.

Osborn, William C., Developing Performance Tests for Training Evaluation, (HumRRO Professional Paper 3-73). Human Resources Research Organization, Alexandria, Virginia, February 1973.

Panitz, A., and Olivo, C., National Occupational Competency Testing Project, Phase I: Planning--Organizing--Pilot Testing. The State of the Art of Occupational Competency Testing (Vol. 2). Office of Education (DHEW), June 1970.

Pickering, Edward J., and Anderson, Adolph V. Measurement of Job-Performance Capabilities, Navy Personnel Research and Development Center, San Diego, California, December 1976.

Swezey, R., and Pearlstein, R., Developing Criterion-Referenced Tests. Applied Science Associates, Inc., Reston, Virginia, for the US Army Research Institute for the Behavioral and Social Sciences, Alexandria, Virginia, September 1974.

Vineberg, Robert and Taylor, Elaine N. Performance in Four Army Jobs by Men at Different Aptitude (AFQT) Levels: 4. Relationships between Performance Criteria, HumRRO Technical Report 72-23. Human Resources Research Organization, Alexandria, Virginia, August 1972.

Appendix A

TASK SELECTION PROCEDURES

APPENDIX A

TASK SELECTION PROCEDURES

INTRODUCTION

The ideal task for use as a test candidate can be defined as one which requires the application of every key and essential behavior component of tasks within the MOS. Such a task does not exist and if one was created solely for test purposes it would lack reality and continuity as it would likely differ from "real world" job performance. Existing tasks and task clusters must then be examined with the goal of identifying those which most closely approximate the "ideal". Candidate tasks then become those which require the application of the greatest number of separate, distinct, key, and essential behaviors which are common to the majority of tasks within the MOS.

DEVELOPMENT OF BEHAVIOR DESIGNATORS

To select the candidate tasks at a given skill level, behavior designators (explained later) are used to identify elements, or performance steps, within each task. A matrix is then developed to identify common elements (behaviors) which cut across tasks and equipment. An example of this matrix is included as Figure A-1, and should be studied as the analyst reads the explanation that begins with Step 1. The matrix is constructed by following the algorithm presented as Figure A-2. In this algorithm, action steps are enclosed by a circle ◯ , questions by a diamond ◇ , and answers by a square □ . Each step is explained in detail following this introduction.

In general the matrix contains two types of entries; the first is a listing of all critical tasks at a given skill level and second is a listing of the behavior designators pertinent to the tasks. This will then enable you to rank order the tasks as candidates for inclusion in an SQT.

TASK SELECTION PROCESS

(1) Identify skill level. Separate tasks by skill level so that only one level will be considered for any one matrix.

(2) Identify critical tasks. Prior to establishing commonality, the importance of each task must be assessed. Most of the time, the documentation which contains the task performance steps will also indicate the criticality or level of importance of the task. If it does not, at least two subject matter experts should assess each task using the following basic guidance. "Two major classes of importance are: (1) criticality to mission accomplishment, based on expert judgments, and (2) performance deficiencies in the field, documented by field data demonstrating weak performance. Potential sources of data include Army Training & Evaluation Program (ARTEP) results, Maintenance Management Center (MMC) data, Equipment Serviceability Criteria (ESC) reports, Inspector General inspection reports, and morning reports".[1/]

◇3◇ Should the critical tasks be separated and grouped by functions? Some MOSs will contain so large a number of critical tasks that some way will have to be found to reduce the job to manageable bites of say, 40 to 60 tasks per matrix. One way is to group the tasks functionally; for instance, tasks in the MOS 62F, "Crane Operator," can logically be grouped by "Maintenance" functions and "Operations" functions.

[4] YES. A yes answer simply means that at this point, you should divide the tasks functionally so that you will be able to construct one matrix for each functional area. Using the example of two MOS 62F tasks which are:

(1) Perform operator's maintenance on the crawler crane.

(2) Drive the truck mounted crane between job sites.

The Job Task Summary Sheet (JTSS) for task (1) lists 18 separate performance steps, ranging from inspection and replacement to lubrication. These are clearly preventive maintenance functions as the behavior designators indicate. The JTSS for task (2) lists six separate performance steps such as positioning the boom, retracting outriggers, and starting and stopping the crane. These are clearly operational functions.

1/ Procedures for Validating Skill Qualification Tests, Stephen F.Hirshfeld, Douglas L. Young, & Milton H. Maier, U.S. Army Research Institute for the Behavioral and Social Sciences, June 1976. (Draft)

After all tasks have been grouped by functions, you would proceed to Step (9)

[5] No. Move on to Step <6>.

<6> Should the critical tasks be grouped by systems? This question is asked for much the same reasons as those explained in Step <3>. For example, the soldier's manual for MOS 63C, "Track Vehicle Mechanic" shows some 290 tasks. These tasks however, can be grouped by several major systems, such as, Engine & Ignition System, Cooling System, Fuel System, Electrical System, Suspension System, etc., and a separate commonality matrix may be constructed for each system.

[7] YES. A yes answer means that at this point, you should separate the critical tasks into groups of systems so that one matrix may be constructed for each group.

[8] NO. If this answer is chosen, proceed directly to Step (9).

(9) List all critical tasks across top of matrix. Maintain the original wording of the task as it appears in the soldier's manual when filling out the top portion of the matrix. This helps to eliminate confusion later. It also helps because the objective of this matrix is <u>not</u> to redefine task statements. By this time you should be able to consider the task statement as <u>valid</u>. All you are required to do at this point is list the tasks selected in Step (2) across the top of the matrix.

(10) Select behavior designators and list them vertically to form the left-hand column of the matrix. Behavior designators are those verbs which denote specific skills or knowledge necessary for task element accomplishment. Examples are shown in the left-hand column of the sample matrix (Figure A-1). The selection of behavior designators is accomplished through a search of the JTSS (Table A-1) Task Data Cards (TDC), or other similar documentation which details the actual performance steps, or task elements, for each critical task.

While the concept of this matrix is applicable to the whole field of MOSs, the behavior designators selected will be considered as unique to the set of tasks being analyzed. This is because the same verb may be used to designate different behaviors in different MOSs. For example, the verb "oil," when used in describing an element of the task "Oil/Wet concrete forms,"

denotes a quite different action than when used in a vehicular maintenance task. At times, the same verb will be used to describe different behaviors within the same MOS.

For example, the verb "saw" describes one skill when used in reference to wood and another when used in reference to concrete. In this situation, the analyst would simply include the modifier as part of the designator, such as saw (wood) or saw (concrete). For example, the task, "Cut and install batter boards" appears on the JTSS as shown in Table A-1. The appropriate designators are underlined.

Remember, you are looking for VERBS, words that describe actions, something the soldier must do in order to accomplish the task. Many times the same verb will be used in each of a dozen performance steps of a single task. An example of this is found in the task "Identify construction material by type and size." That's fine; the verb "identify" is describing basically the same action each time. Whether the soldier must identify nails or grades of lumber, it is still basically the same action. Simply write the word "identify" in the left-hand column and go on to the next designator or to the next JTSS if there are no more different designators in that task.

(11) Analyze the JTSS for each task and plot the designators by checking them off on the matrix as they apply to each task. You should begin with the JTSS for the first task you have listed at the top of the matrix as has been done in Figure A-1.

Each behavior designator is given equal weight. Thus, only one check (or point) would be given per identified behavior per task so that although a single task may contain many performance steps where certain behaviors occur more times than others, none would be weighted more heavily than any other.

Notice that although the task in the JTSS lists two separate cutting actions, once for the posts and once again for the boards themselves, the designator "cut" would receive only one point. This insures that each action receives the same point value or "weight" in the matrix.

(12) Sum designators. Once the matrix is complete, that is, after all critical tasks have been accounted for, sum the behavior designators horizontally across each task. This step establishes

TABLE A-1. JOB TASK SUMMARY SHEET

Task: Cut and install batter boards	Task Criticality (Circle One) (C) I N *	
Steps in Performance	Standard of Performance	Materials, Tools, Equipment
1. Cut 12 batter board posts and sharpen one end	Posts will be cut long enough so that when driven firmly into the ground the posts will extend above required finish elevation of the foundations as directed by crew chief.	2 x 4 material, 6-ft folding rule, square, crosscut saw, half hatchet
2. Emplace batter board post at corners	3 batter board posts will be firmly driven into the ground 3 or 4 feet outside of each corner post as directed by crew chief.	12 - 2 x 4 stakes maul or sledge, folding rule, framing square
3. Measure and cut batter board	Batter boards will be cut long enough to be securely fastened from center post to outside post as directed by crew chief.	1 x 6 material, folding rule, square, crosscut saw
4. Attach batter boards to posts	Batter boards will be securely nailed to the posts, level and at exact elevation of finish foundation as directed by crew chief.	1 x 6 batter boards, claw hammer, folding rule, carpenter's level, 8d common nails

The behavior designators are: Cut, Sharpen, Emplace, Measure and Nail**

* Task Criticality Code

C - Critical
I - Important
N - Not important

** The designator "nail" is used here instead of "attach" as "nail" appeared to be the more definitive designator. "Attach" would probably also be in your list. The important thing to remember is that only one or the other would be checked as checking both designators for the same action would result in an improper weighting of the action.

which designators occur with the greatest frequency across the critical tasks and is the first step towards identifying commonality.

(13) Establish mean. The column formed by these totals (Step (12)) is then summed vertically, and the total is shown in the lower right-hand corner of Figure A-1. This total is then divided by the number of values (or entries) in the column to establish a mean. This gives you the average number of tasks in which a designator occurs.

(14) Having established a mean number of tasks in which a behavior occurs, those behaviors which occur across tasks with a frequency at or above the mean are considered common. You now go through the matrix and circle the check marks of every element that is common.

(15) Evaluate designators. At this point you look for behaviors which are critical even though they may not be common. For example, the designator "vibrate (concrete)" In Figure A-1 is not identified as common. You as a subject matter expert, however, may consider it to be a behavior which is essential to mastery at this skill level. You would therefore circle the check marks applying to that designator so that tasks which incorporate it are given an extra "weight" which will result in the task being ranked higher in Step (17) . Remember, the matrix is a tool to aid in task selection and test development; as such, it should not become an absolute basis for the selection/rejection of test item candidates. The following criteria are given as a guide to evaluating behavior elements for importance and criticality.

(1) The degree of skill required in the use of tools, equipment, or communication - the higher the degree, the more critical the element.

(2) The time required to master the skill - the more time, the more critical.

(3) Frequency of performance of the skill - the more frequent, the more critical.

(4) Consequences of failure to perform - jeopardy to life and equipment equals criticality.

(5) Degree and caliber of reaction required - unfailing, rapid performance under all conditions equals criticality.

(16) Vertically sum all circled (common and/or critical) designators under each task. This enables you to establish which tasks are the prime candidates for inclusion in an SQT.

(17) Rank order tasks. With common behavioral elements preliminarily identified, preliminary candidate tasks for inclusion in the SQT are rank ordered according to the number of separate common behavioral elements each contains. Thus, the task or tasks with the greatest number of circled check marks would become the first task selected for scanning in the Simulation Procedures.

(9) – Critical ta... the soldie... MOS & ...

(10) – Derived from a thorough search of job task analysis.

BEHAVIOR DESIGNATORS (23 designators)	CRITICAL TASKS									
	Identify construction material by type & size	Cut & install batter boards	Construct, replace footers & columns	Frame walls & partitions	Assemble roof trusses using template	Install roof trusses	Place & finish concrete	Install anchor bolts in concrete	Construct joints in concrete	Prepare timber ... for driving
Identify	X									
Check			(X)	(X)	(X)			(X)		(X)
Measure		(X)	(X)	(X)	(X)	(X)				
Square		(X)		(X)	(X)					(X)
Plumb				X		X				
Cut		(X)	(X)	(X)	(X)				(X)	(X)
Sharpen		X								
Emplace		X								
Nail		(X)	(X)	(X)	(X)	(X)				
Collect			(X)	(X)				(X)		(X)
Excavate			X							
Mix			X							
Place (Concrete)			(X)				(X)		(X)	
Trowel			(X)				(X)		(X)	
Raise w/jacks			X							
Bore								X	X	
Saw (Concrete)									X	
Screed			X				X			
Edge			X				X			
Broom (Concrete)										
Vibrate (Concrete)			X				X			
Wood Float Finish			X				X			
Lay block/brick			X							
Common Behavior Totals	0	4	7	6	5	2	2	2	3	4

(16) – Sum of all circled designators.

(17) – Rank order of the tasks based on the greatest number of circled checkmarks.

(9) – Critical task extracted verbatim from the soldier's manual supporting this MOS & skill level.

Place & finish concrete	Install anchor bolts in concrete	Construct joints in concrete	Prepare timber piles for driving	Frequency behavior designator appears in the critical tasks. See Notes 1 & 2
				1
	(X)		(X)	(5)
				(5)
			(X)	(4)
				2
		(X)	(X)	(6)
				1
				1
				(5)
	(X)		(X)	(4)
				1
				1
(X)		(X)		(3)
(X)		(X)		(3)
				1
	X	X		2
		X		1
X				2
X				2
				0
X				2
X				2
				1
2	2	3	4	55

Rank order of the tasks based on the greatest number of circled checkmarks.

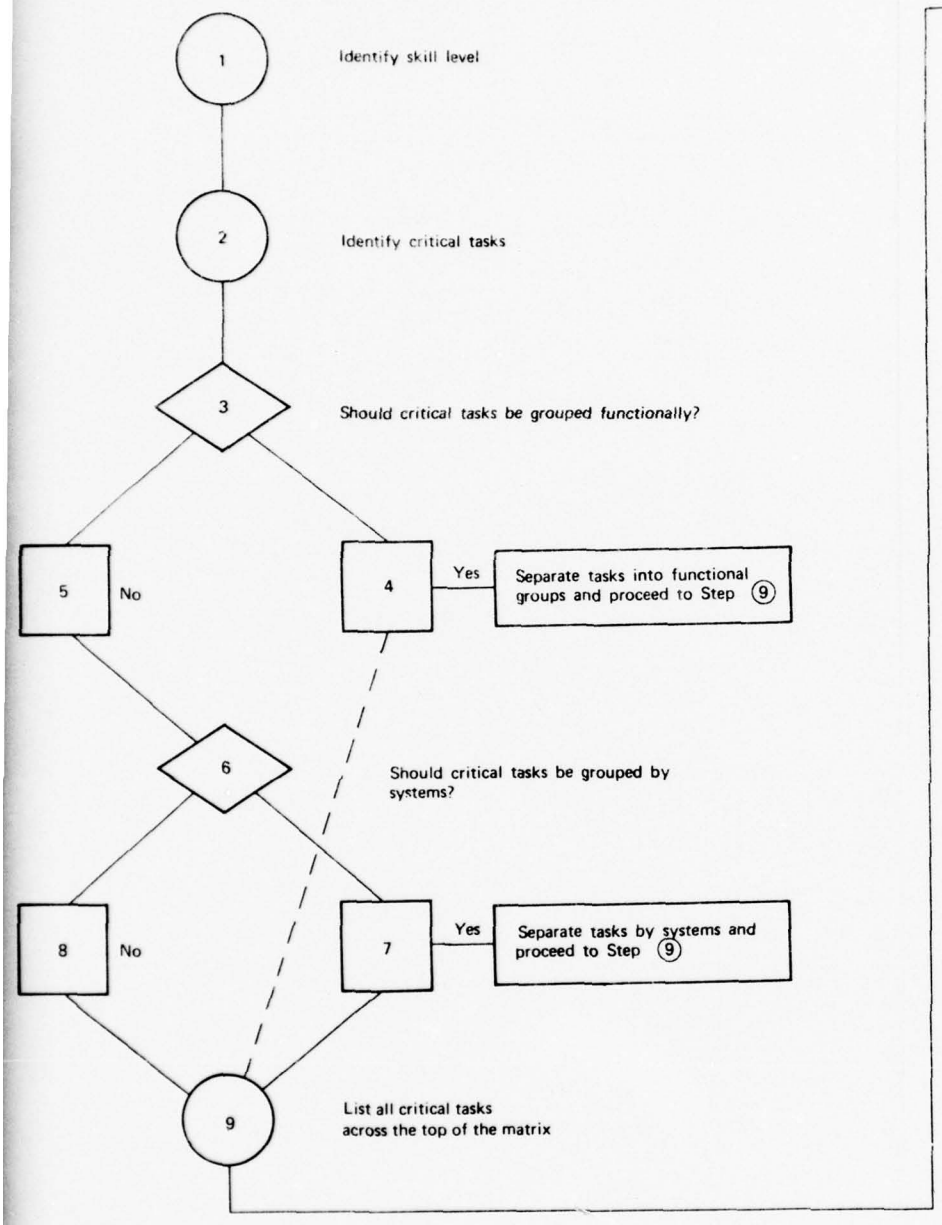
NOTES:
1. ( ) numbers indicate the sum of those designators which occur more than the established mean (2.4) & are further identified by (X). Detailed explanation contained in Step (14) of the text.
2. Numbers circled e.g., (10) correspond to the appropriate step in the algorithm & are explained in detail in the text of the handbook.

Total number of behavioral designators.

÷ 23 = 2.4

Average number of tasks in which a behavior designator appears.

Figure A-1. Task Selection Matrix



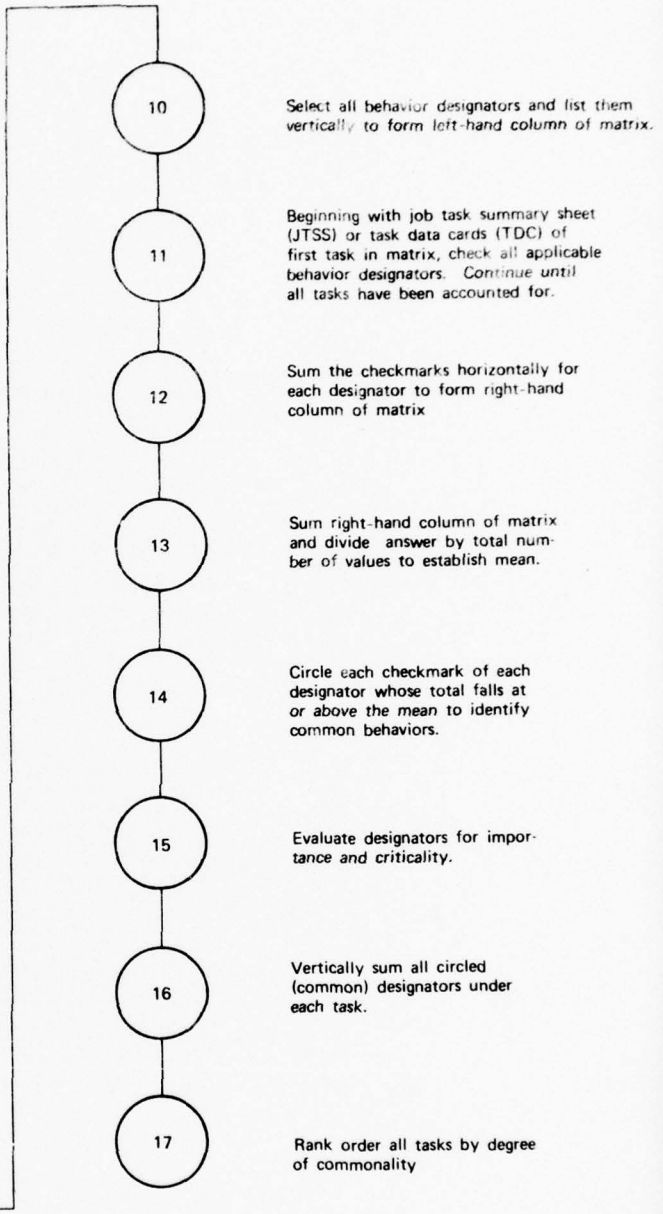


Figure A-2. Task Selection Procedures

Appendix B

SIMULATION PROCEDURES

APPENDIX B

SIMULATION PROCEDURES

INTRODUCTION

The procedures are presented as an algorithm (see Figure B-1) made up of a series of actions, questions, answers, and decisions involved in the development of the audio-visual performance test. Segments of the algorithm are displayed at the end of each section for easy reference. Actions are indicated by a circle, ◯; questions by a diamond, ◇; answers by a square □,; and discussions are enclosed within a rectangle, ▭. The procedures are easy to follow. It has 54 steps, and an explanation is provided for each step. In Steps 1 through 33 actual task element data is considered to provide a partial demonstration of how the procedures are used.

PRELIMINARY TEST MODE SELECTION

GENERAL

The tasks selected for testing must be analyzed to determine whether or not a realistic, reliable, and valid scoreable unit can be presented in an audio-visual (A/V) test mode. Your earlier analysis in the Task Selection Procedures provided a rank ordered list of tasks to be considered in developing the SQT. In selecting the mode by which the tasks may be tested, one test mode may be more appropriate than another for a specific task. A hands-on test may be the most appropriate for one task, while a written or A/V mode would be appropriate for another.

This section of the procedures is primarily designed to aid in making a preliminary determination of whether a task is suitable for testing by A/V. However, in the process it becomes necessary to identify whether a performance or written test is appropriate.

PRECEDING PAGE BLANK

Figure B-2 depicts Steps 1 through 11 and shows the sequence of operations involved in making a preliminary selection of test mode.

SIMULATION ALGORITHM

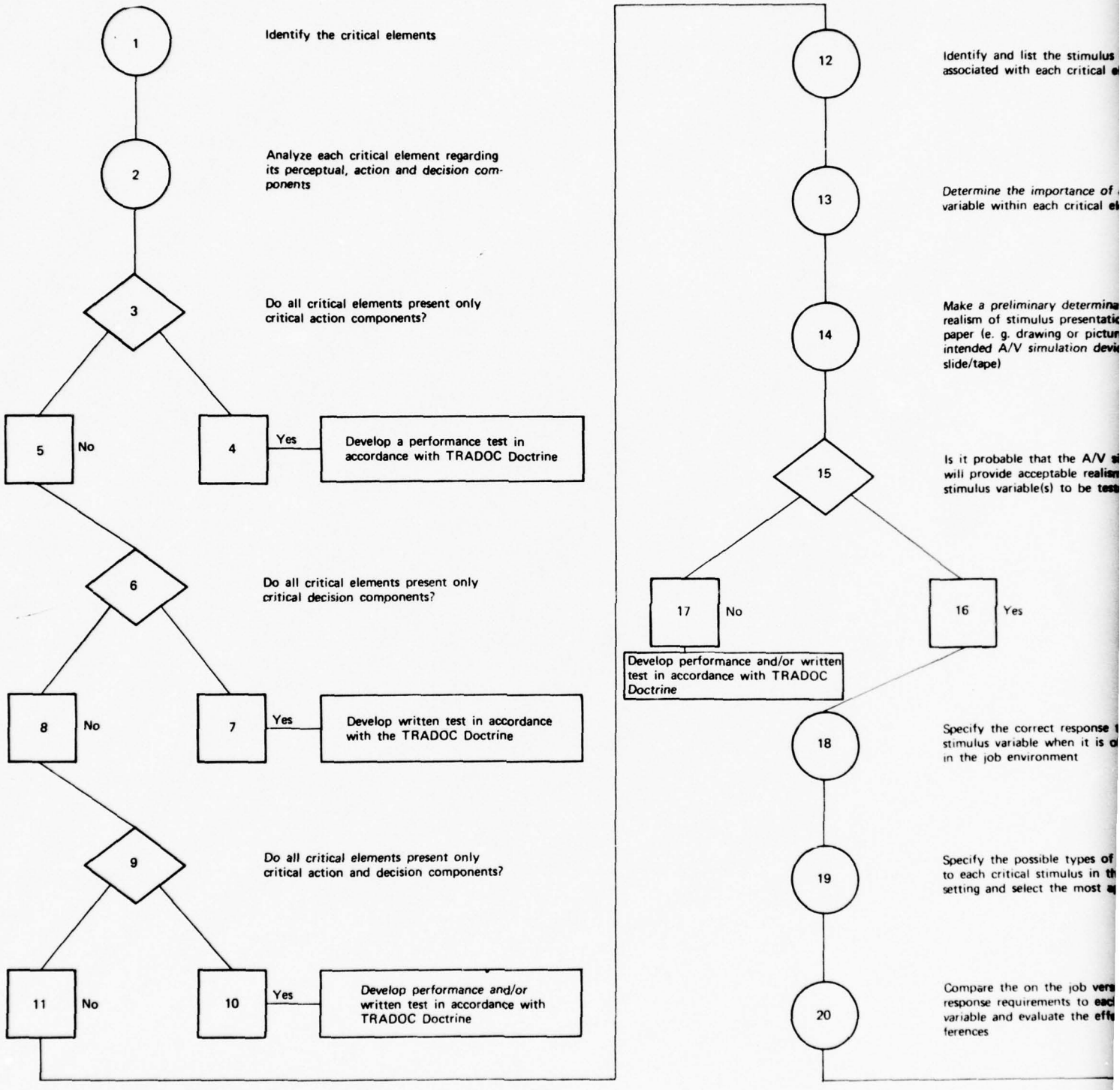
(1) Identify the critical elements. Each task being considered for audio-visual simulation testing has been identified as a candidate task to be covered within the SQT. The purpose of this first step is to identify the parts of the task which need to be tested. Each task includes a number of steps, or elements, which are listed on the JTSS/TDCs or in the soldier's manual. Some of the elements must be accomplished with a high degree of accuracy if the task is to be finished in an acceptable manner; other elements must be accomplished, but some error can be tolerated without serious effect on the performance quality of the task. The elements that must be accomplished with a high degree of accuracy are the critical or key elements, and therefore should be selected for testing. For some tasks all elements may be critical, but for others, all may not.

The determination of critical elements must be made by personnel who are skilled and experienced in the task (i.e., subject matter experts). Ideally, at least three subject matter experts should be involved in this process. The end product of this step is a list of the critical elements associated with a given task.

EXAMPLE - Throughout this explanation, we will use the MOS 51B task, "Direct/Control Placing and Finishing Concrete" as our example. It will be assumed that all elements of this task are critical. These are listed in Column 1 of Table B-1.

(2) Analyze each task element regarding its perceptual, action, and decision components. This step will help you determine the best test mode for each of the critical elements listed in the element analysis table by assessing the importance of three component activities in each element. The three component activities are perceptual, action, and decision.

Perceptual - That component of a task element which involves judgments based upon the senses (see, hear, touch, taste, smell).


1
Identify the critical elements
2
Analyze each critical element regarding its perceptual, action and decision components
3
Do all critical elements present only critical action components?
5
No
4
Yes
Develop a performance test in accordance with TRADOC Doctrine
6
Do all critical elements present only critical decision components?
8
No
7
Yes
Develop written test in accordance with the TRADOC Doctrine
9
Do all critical elements present only critical action and decision components?
11
No
10
Yes
Develop performance and/or written test in accordance with TRADOC Doctrine
12
Identify and list the stimulus
associated with each critical e
13
Determine the importance of
variable within each critical e
14
Make a preliminary determina
realism of stimulus presentatio
paper (e. g. drawing or pictur
intended A/V simulation devi
slide/tape)
15
Is it probable that the A/V si
will provide acceptable realism
stimulus variable(s) to be test
17
No
16
Yes
Develop performance and/or written test in accordance with TRADOC Doctrine
18
Specify the correct response t
stimulus variable when it is o
in the job environment
19
Specify the possible types of
to each critical stimulus in th
setting and select the most a
20
Compare the on the job vers
response requirements to eac
variable and evaluate the effe
ferences

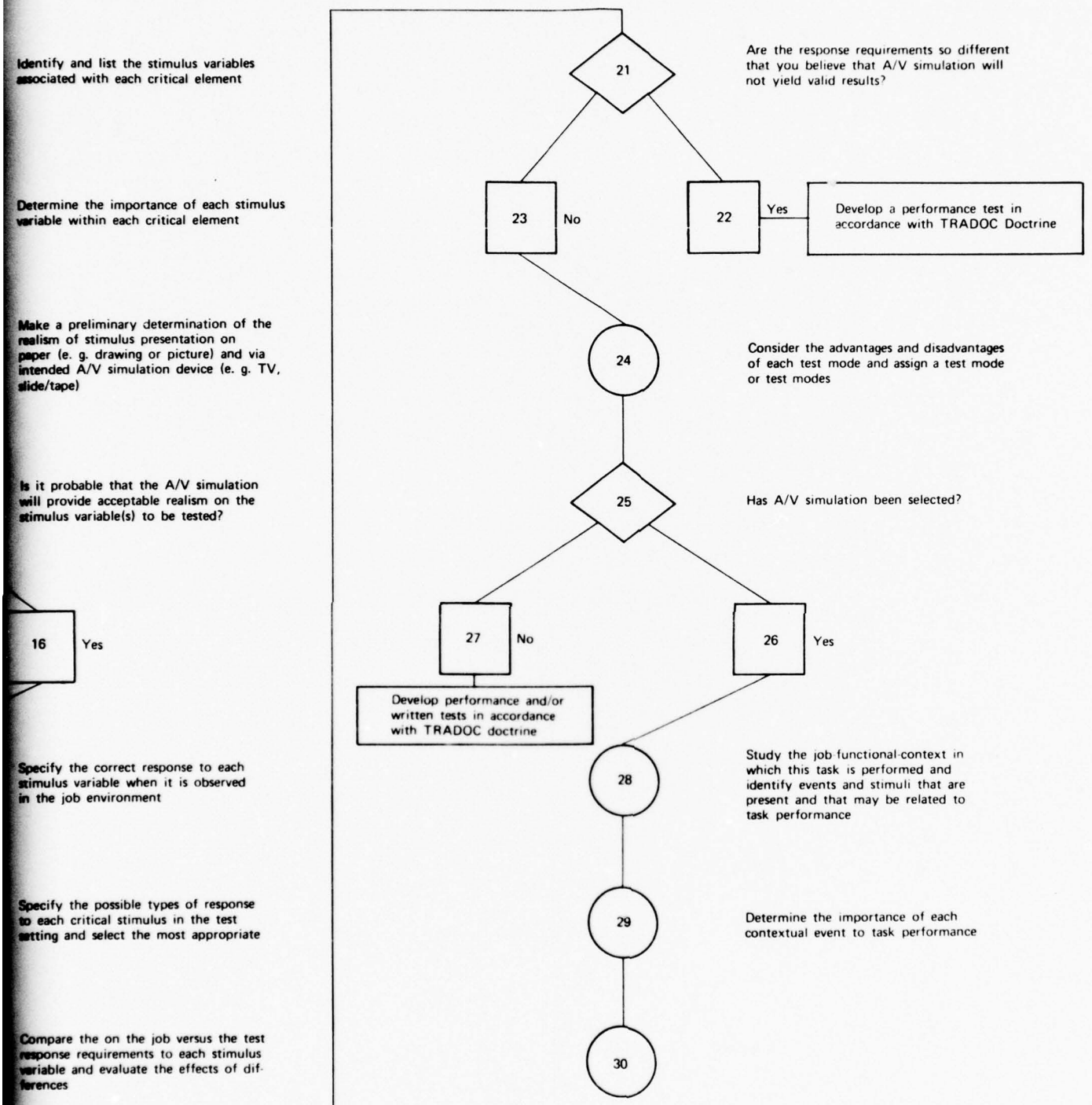
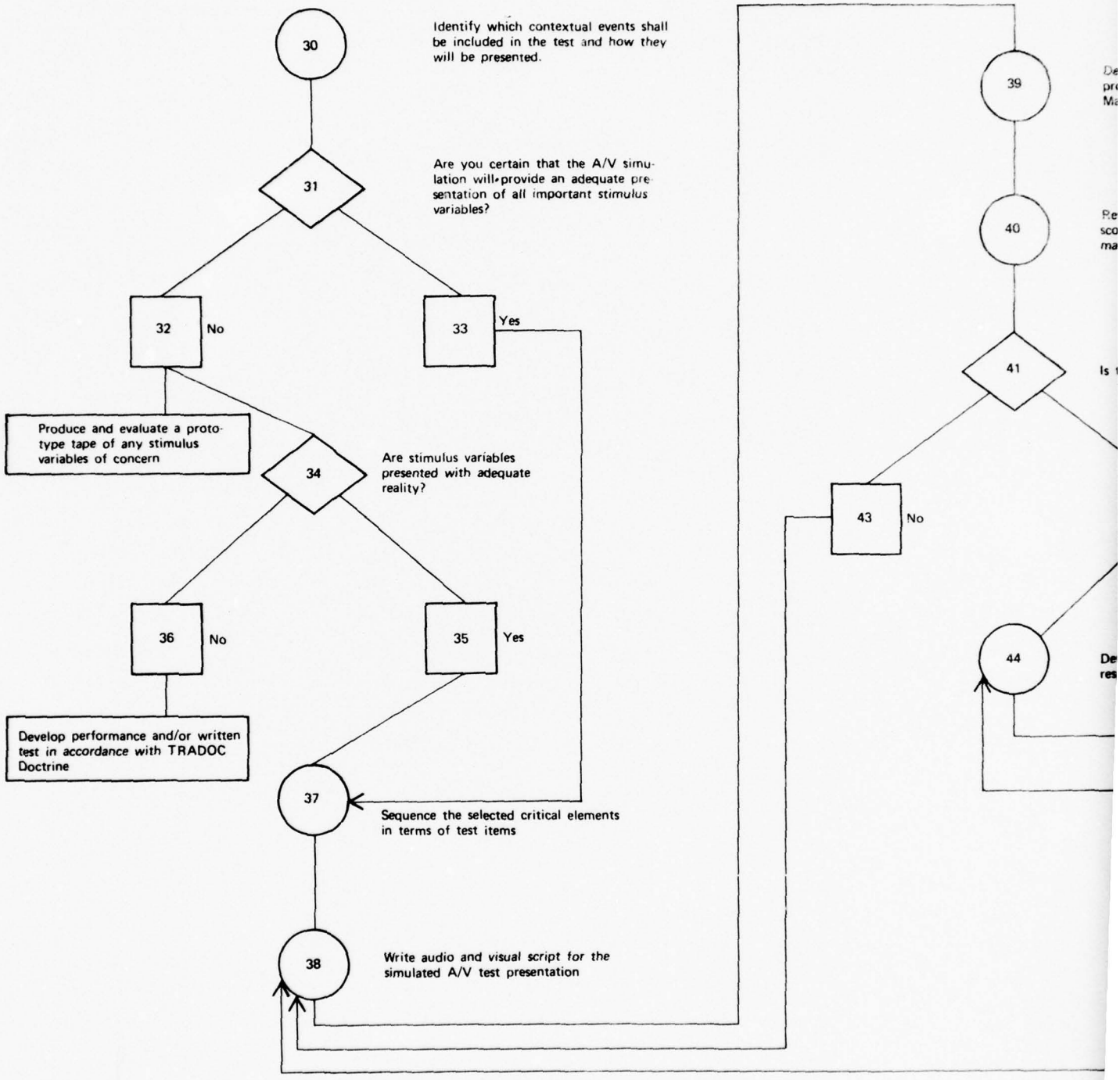


Figure B-1. Simulation Algorithm (Sheet 1 of 2)


30
Identify which contextual events shall be included in the test and how they will be presented.
31
Are you certain that the A/V simulation will provide an adequate presentation of all important stimulus variables?
32
No
33
Yes
Produce and evaluate a prototype tape of any stimulus variables of concern
34
Are stimulus variables presented with adequate reality?
36
No
35
Yes
Develop performance and/or written test in accordance with TRADOC Doctrine
37
Sequence the selected critical elements in terms of test items
38
Write audio and visual script for the simulated A/V test presentation
39
40
41
43
No
44

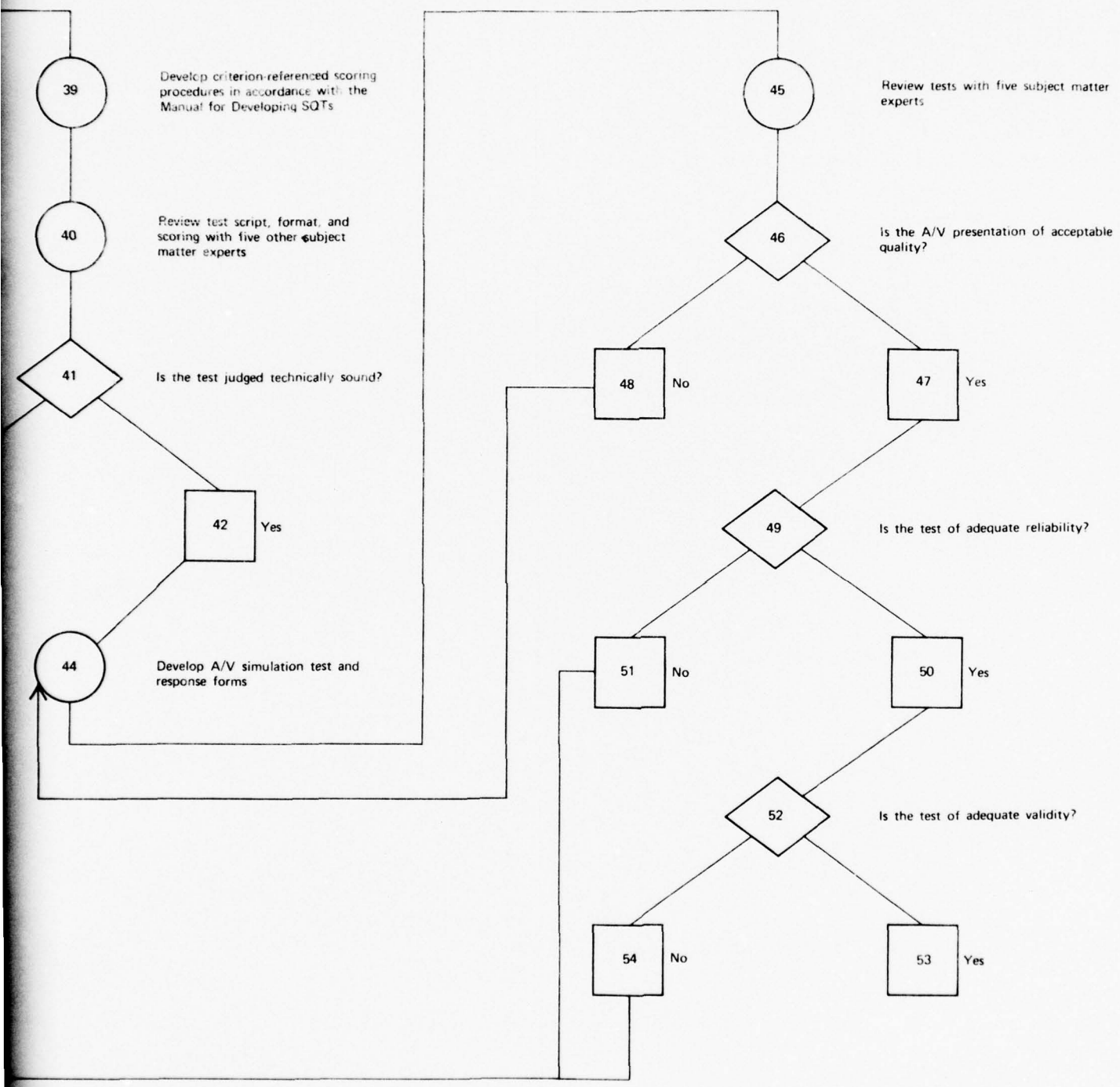


Figure B-1. Simulation Algorithm (Sheet 2 of 2)

2

TABLE B-1. ELEMENT ANALYSIS (A)

Example Analysis of Critical Elements for Task "Direct/Control Placing and Finishing Concrete"

Critical Elements	Critical Components	Stimulus Variables
Column 1	Column 2	Column 3
1. Direct/Control placing of ramps		
2. Direct/Control placing of concrete for slab construction or small paved surface on grade		
3. Direct/Control placing concrete into wall, beams, and girder forms		
4. Direct/Control use of vibrator		
5. Direct/Control screeding of concrete		
6. Direct/Control finishing concrete using a wood float		
7. Use long-handle wood float		
8. Direct/Control finishing concrete using steel finishing trowel		

PRECEDING PAGE BLANK

Action - That component of a task element involving bodily movement (motor skills).

Decision - That part of a task element which involves using past knowledge and new information to determine when or how to perform the task element.

Most all tasks contain perceptual, action, and decision components. To identify these components in a task element, a subject matter expert should consider the following three questions:

(1) When errors occur in the performance of this task element, is it because people fail to perceive (see, hear, feel, taste, smell) important information?

If the answer to the question is "YES," the task element has a critical perceptual component.

(2) When errors occur in the performance of this task element, is it because people fail to make coordinated or precise bodily movements?

If the answer to the question is "YES," the task element has a critical action component.

(3) When errors occur in the performance of this task element is it because people are misusing knowledge or information?

If the answer to the question is "YES," the task element has a critical decision component.

EXAMPLE - The critical components for the elements of the task to "Direct/Control Placing and Finishing Concrete" are listed in Column 2 of Table B-2. You should note that:

(1) Only one task shows a critical action component, all the other elements require directing others rather than doing the act.

(2) The first three elements contain only critical decision components because: (a) it is considered likely that errors in those elements would result from failure to give proper direction even though one perceived the situation accurately, and (b) the elements do not involve the actual placing of ramps or concrete.

TABLE B-2. ELEMENT ANALYSIS (B)

Example Analysis of Critical Elements for Task: "Direct/Control Placing and Finishing Concrete"

Critical Elements	Critical Components	Stimulus Variables
Column 1	Column 2	Column 3
1. Direct/Control placing of ramps	Decision	
2. Direct/Control placing of concrete for slab construction or small paved surfaces on grade	Decision	
3. Direct/Control placing concrete into wall, beams, and girder forms	Decision	
4. Direct/Control use of vibrator	Perceptual/Decision	
5. Direct/Control screeding of concrete	Perceptual/Decision	
6. Direct/Control finishing concrete using a wood float	Perceptual/Decision	
7. Use long-handle wood float	Action/Perceptual	
8. Direct/Control finishing concrete using steel finishing trowel	Perceptual/Decision	

(3) Elements 4, 5, 6, and 8 contain critical perceptual and decision components because errors are more likely to occur when either: (a) the person fails to recognize the consistency, wetness, or level of the concrete relative to the operations that must be performed, or (b) the person recognizes the consistency, wetness, or level but directs an operation to begin or end at the wrong time.

◇3◇ Do all critical elements present only action components? This question is asked to determine whether the task should be tested by a performance test or considered for testing via written or audio-visual simulation mode. This question is easily answered by referring to your element analysis table.

[4] YES. If this answer is selected, it leads to the recommendation that a performance test be developed for the task. The assumption is that the action component of a task element is best measured via a performance test, and that when all elements present only critical action components, a performance test is especially justified.

[5] NO. If this answer is selected, it is necessary to gain more information, which is done by moving on to Step ◇6◇ .

EXAMPLE - Column 2 of Table B-2 shows that there were critical decision or perceptual components for at least one task element. Therefore, the answer to the question presented in Step ◇3◇ is "NO." A "YES" answer would require that each task element has only a critical action component.

◇6◇ Do all critical elements present only critical decision components? The answer to this question determines whether the task should be tested by a written test or if an audio-visual test format can be used. The question is easily answered by referring to your element analysis table.

[7] YES. If this answer is selected, it leads to the recommendation that a written test be developed for the task. The assumption is that when the decision component is the only critical component, it can be measured most economically via a written test.

[8] NO. If this answer is selected, it is necessary to gain more information which is done by moving on to Step ◇9◇

EXAMPLE - Table B-2 shows that there were critical decision or perceptual components for at least one element. Therefore, the answer to Step ◇6 is "NO." If each task element had only a critical decision component, the answer would have been "YES."

◇9 Do all critical elements present only critical action and decision components? This question is asked to determine whether a written and/or performance test should be used or if an audio-visual format should be considered. This question is easily answered by referring to the element analysis table.

[10] YES. If this answer is selected, it is recommended that a written and/or performance test be developed for the task. Typically, a performance test is preferred. However, sometimes the decision component of a task or subtask can be tested adequately via performance testing only by many repetitions of the action components. This can be impractical. For example, a simple performance test of one's ability to construct a given type of roof may adequately measure the action components (measure, saw, nail) common to many types of roofs, but it may not adequately test the decision components (selection of materials, determining sequence of events) that vary according to the type of roof. In this situation a written test, which addresses the decision components, might be used along with the performance test which focuses on the action components.

[11] NO. If this answer is selected, then by process of elimination, you have isolated those task elements which may feasibly be tested in an A/V mode. It is now necessary to gain more information which is done by moving to Step (12)

EXAMPLE - Column 2 of Table B-2 shows that there are perceptual components as well as action and decision components. Therefore, the answer to this step is "NO." If only action and decision components were listed, the answer would have been "YES."

TEST REALISM

GENERAL

Since you have determined that a task/element has perceptual content and is a candidate for A/V testing, further analysis becomes necessary. To

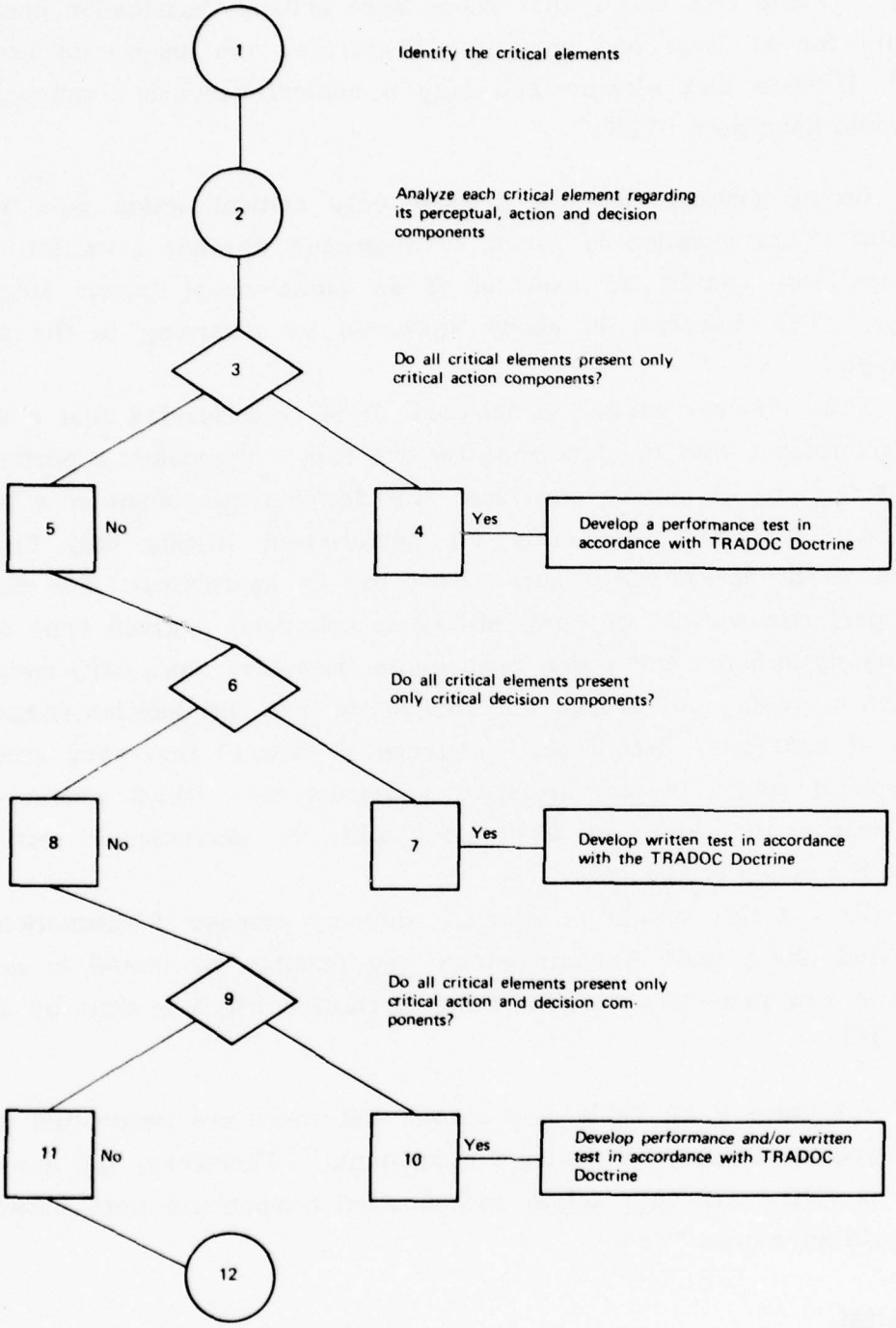


Figure B-2. Preliminary Test Mode Selection

provide a valid test, the A/V mode must present the task/element in a realistic manner. Figure B-3 (Steps 12 through 17), at the end of this section, outlines the sequence of operations leading to this determination.

SIMULATION ALGORITHM

(12) Identify and list the stimulus variable associated with each critical element. The purpose of this step is to analyze each task element and specifically determine what a person responds to as he performs the task element. For example, in deciding when to use a wood float for finishing concrete, it is not adequate that a person respond only to the "appearance" of the concrete or to the amount of moisture in the concrete. There are certain clues, or stimulus variables that permit a person to judge the appearance or moisture of the concrete, and each of these stimulus variables must be listed.

EXAMPLE - In Column 3 of Table B-3, five possible stimulus variables have been listed for the sixth critical step - "Direct/Control finishing concrete using a wood float." This example indicates that in directing the finishing of concrete using a wood float, a person responds to certain characteristics of the appearance of the concrete; specifically, he determines when to start and stop based upon the five stimulus variables listed in Column 3.

(13) Determine the importance of each stimulus variable within each critical element. The purpose of this step is to further define the stimulation requirements of a given task or task element. In performing this step, subject matter experts should refer to the stimulus variables which were identified in Step (12) and evaluate their importance to the proper performance of the critical element. Simulation usually degrades some aspects of the stimulus. If this step is performed properly, it helps anticipate the effect of any degradation.

Importance is difficult to determine because many factors go into making something important. In this step a procedure is suggested; but it is recognized that the subject matter experts will have to be subjective. The subject matter expert should answer the following question in determining the importance of a stimulus variable. "Considering all of the information which is provided by all of the stimulus variables which are typically present when this task element is performed, how often does this one stimulus variable provide unique information that is essential to proper performance of the task element?"

TABLE B-3. ELEMENT ANALYSIS (C)

Example Analysis of Critical Elements for Task: "Direct/Control Placing and Finishing Concrete"

Critical Element	Critical Components	Stimulus Variables
Column 1	Column 2	Column 3
1. Direct/Control placing of ramps	Decision	
2. Direct/Control placing of concrete for slab construction or small paved surface on grade	Decision	
3. Direct/Control placing concrete into wall, beams, and girder forms	Decision	
4. Direct/Control use of vibrator	Perceptual/Decision	
5. Direct/Control screeding of concrete using a wood float	Perceptual/Decision	
6. Direct/Control finishing concrete using a wood float	Perceptual/Decision	a. Uniformity of color of concrete b. Presence or absence of swirls in concrete c. Presence or absence of pits in concrete d. Presence or absence of pockets of water e. Firmness of concrete in response to slight pressure
7. Use long handle wood float	Action/Perceptual	
8. Direct/Control finishing concrete using steel finishing trowel	Perceptual/Decision	

Obviously, a stimulus variable is very important if it always provides unique information (i.e., information that is not available from any other stimulus variable) which is also essential to proper performance. A stimulus characteristic will be much less important if it either duplicates information already available or if its informational value has a less direct effect upon proper performance. Each stimulus variable should be labeled as very important, moderately important, or not very important.

EXAMPLE - In performing this step, the subject matter experts will review the stimulus variables associated with a task element. See Column 3 of Table B-3 and note the five stimulus variables associated with the sixth critical element. A judgment is now made of the importance of each stimulus variable. The subject matter expert should follow the procedure shown in Table B-4. In that table, each stimulus variable is rated as providing either unique or non-unique information, and as providing information which is essential or non-essential to the outcome of the critical element.

In the example provided in Table B-4, a stimulus variable is rated: (1) very important if it was judged both unique and essential, and (2) moderately important if it is either unique or essential, but not both. A stimulus variable would be rated as not very important if it is neither unique nor essential.

(14) Make a preliminary determination of the realism of stimulus presentation on paper, e.g., drawing or picture and via intended A/V simulation device, e.g., TV, slide/tape. The purpose of this step is to start determining whether a critical element is best tested via an A/V device or a paper-and-pencil format. Both the subject matter expert and a training media expert must work together. They will look at each stimulus variable selected in Step (12) and determine which will be the most realistic method of presentation: (1) a drawing or picture on paper, or (2) an audio-visual device (e.g., TV, slide/tape). In many cases there will be little or no difference, but there are at least three situations in which A/V can add to the realism of a stimulus presentation. First, A/V is advantageous when the observation of motion or of constantly changing physical characteristics of an environment is important. Second, three-dimensional relationships can be presented more effectively via A/V. Finally, the coordination of sound and visual stimuli is most effectively accomplished via A/V.

TABLE B-4. DETERMINING THE IMPORTANCE OF EACH STIMULUS VARIABLE ASSOCIATED WITH A CRITICAL ELEMENT

Critical Element: Direct/Control Finishing Concrete Using a Wood Float

Stimulus Variables	Unique *	Essential *	Importance *
Uniformity of color of concrete	Yes	No	Moderately Important
Presence or absence of swirls in concrete	Yes	Yes	Very Important
Presence or absence of pits in concrete	Yes	Yes	Very Important
Presence or absence of pockets of water	Yes	Yes	Very Important
Firmness of concrete in response to slight pressure	Yes	Yes	Very Important

*Ratings are for purpose of example and are not necessarily valid in describing the variable.

There are also conditions which would favor a drawing or picture. Small color differences will be presented more faithfully with photographs than TV, especially when TV testing involves presentations over many TV sets which are in varying states of repair and adjustment. Likewise when fine-line definition is required, the drawing or photograph will often be preferred.

EXAMPLE - Five stimulus variables from Column 3 of the element analysis table are listed again in the left-hand column of Table B-5. An estimate is now made of whether a drawing or picture versus an A/V presentation of the stimulus variables will provide greater realism. In this situation the A/V media is television, so the realism of a televised presentation is considered.

TABLE B-5. DETERMINING THE APPROPRIATENESS OF AN AUDIO–VISUAL PRESENTATION MODE FOR A SIMULATED SKILL QUALIFICATION TEST

Critical Element: Direct/Control Finishing Concrete Using a Wood Float

(1) Stimulus Variables	(2) Importance	(3) Realism	(4) Recommended Presentation Format
Uniformity of color of concrete	Moderately important	*Still picture acceptable - TV questionable	Still picture (if tested)
Presence or absence of swirls in concrete	Very important	TV acceptable Still picture acceptable	TV
Presence or absence of pits in concrete	Very important	TV acceptable Still picture acceptable	TV
Presence or absence of pockets of water	Very important	TV acceptable Still picture acceptable	TV
Firmness of concrete in response to slight pressure	Very important	TV acceptable Still picture acceptable	TV

*Preferred presentation format

Sometimes it will be necessary to produce a test item in two or more formats to determine the most realistic format, but at this time an estimate is made based on previous experience. Estimates of the most acceptable method of presentation for each of the stimulus variables are listed in the fourth column of Table B-5. The reasons for these estimates are:

(1) A still picture is preferred for the first stimulus variable because it is anticipated that it will be difficult to maintain a constant presentation of small color differences of concrete when shown on different TV sets.

(2) A televised presentation is preferred for the second, third, and fourth stimulus variables because these variables are typically observed as one scans and moves around the perimeter of the concrete. Television will provide for greater realism in simulating the behavior which occurs as these stimulus variables are observed.

<15> Is it probable that the A/V simulation will provide acceptable realism in presenting very important stimulus variables? One may wonder why fourteen steps precede such an obvious question. The rationale up to this point is that: (1) the presence of perceptual content should be established before considering the use of A/V, and (2) the specific simulation requirements, which are the stimulus variables, must be identified before rejecting or accepting an A/V mode. This step is to confirm a preliminary determination of whether or not to use A/V. The answer to this step is based upon the entries in Columns 2 and 3 of Table B-5. Column 2 contains the judgments made earlier in Step (13) and Column 3 contains the judgments made in Step (14) .

[16] YES. This answer indicates the potential use of an A/V format. However, further analysis is required.

[17] NO. This answer indicates that none of the important stimulus variables can be presented by A/V with acceptable realism. Consequently, performance or written testing is necessary.

EXAMPLE - The entries in Columns 2 and 3 of Table B-5 indicate that four of the five stimulus variables judged to be very important and TV is the preferred format for each of these. Consequently, a "YES" answer is clearly indicated. If acceptable realism was possible on only one, two, or three of the variables, a "YES" answer should still be given. In a later step (Step (24)), an A/V test format will be rejected if it is judged to be too limited in scope.

RESPONSE REALISM

GENERAL

To determine whether or not the A/V presentation will provide valid test results, it is necessary to determine whether the test responses and job

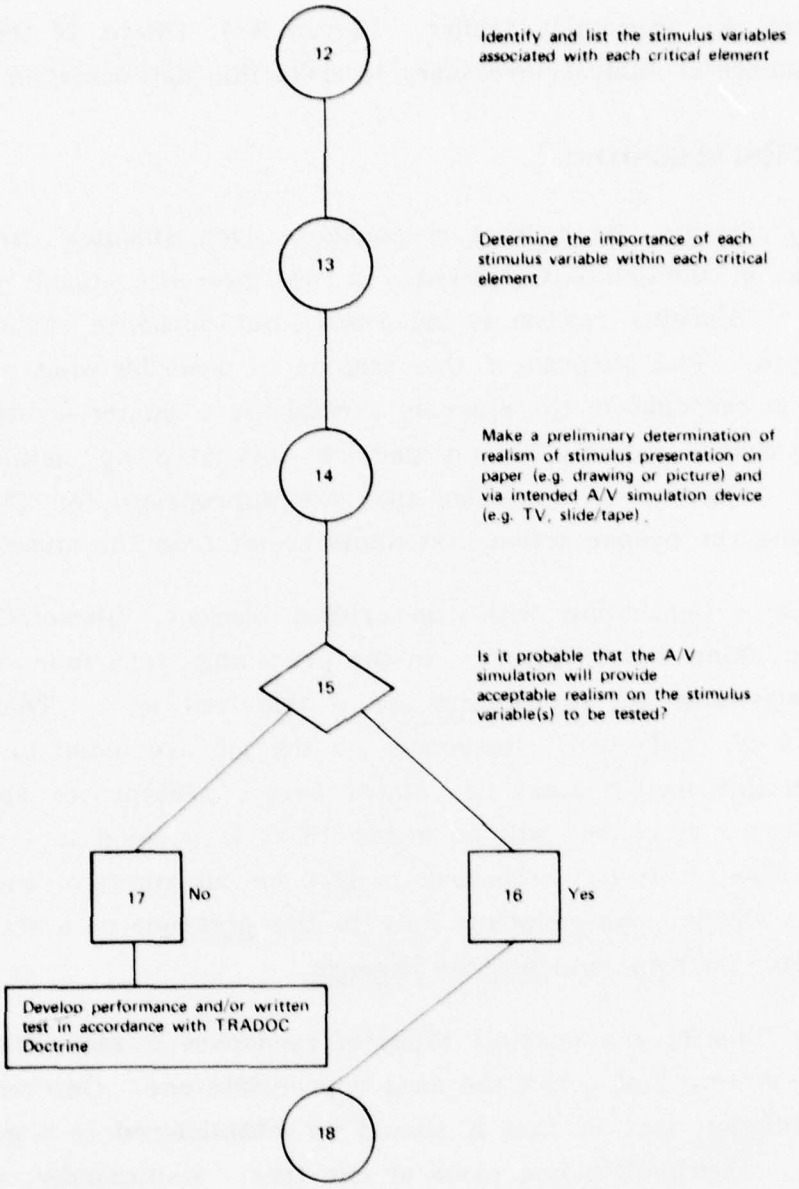


Figure B-3. Test Realism

responses are adequately similar. Figure B-4, (Steps 18 through 23) shows the sequence of analysis necessary to make this determination.

SIMULATION ALGORITHM

(18) Specify the correct response to each stimulus variable when it is observed in the job environment. A test presents stimuli and requires responses. Stimulus realism is important, but response realism must also be considered. The purpose of this step is to describe what a person does on the job in response to the stimulus variable or a cluster of stimulus variables. Subject matter experts should perform this step by listing each stimulus variable or group of variables that are appropriate for TV simulation and identifying the proper action that would result from the stimulus.

EXAMPLE - Continuing with the critical element, "Direct/Control Finishing Concrete Using a Wood Float." In the preceding steps four stimulus variables were designated as appropriate for a televised test. These are listed in Column 1 of Table B-6. Responses on the job are listed in Column 2. Since each variable has at least two states (e.g., presence or absence), at least two different responses will be needed if it is decided to test for all states. In this case, it is not necessary to test for all states of each variable, because on the job one responds only to the presence of a state (i.e., swirls, pits, water pockets) and not the absence.

(19) Specify the possible types of responses to each stimulus variable in the test setting and select the most appropriate one. One requirement of the A/V simulation test is that it should be administered to a group of soldiers who are assembled in one place at one time. Additionally, at this time, the use of computer response terminals or actual equipment is not considered. Consequently, responses will be recorded on paper, and the range of types of responses is limited. However, it is emphasized that there are alternatives to the standard multiple-choice response format, and some of these alternatives will definitely be preferred in certain situations.

The purpose of this step is to select the most appropriate, or job-like response that is available for the A/V simulation test format. The subject matter expert should use his ingenuity in designing or selecting the type of test response. Four categories of test responses are now discussed in some detail.

TABLE B-6. COMPARISON OF RESPONSE REQUIREMENTS TO STIMULUS VARIABLE

(Step ⑫) Stimulus Variables	(Step ⑱) Response on the Job	(Step ⑲) Response possibilities on TV Test	(Step ⑳)	**Recommended Presentation Format
1. Presence or absence of swirls in concrete				
a. If present	a. Direct worker to remove swirl	*a.(1) Report observing a swirl on answer sheet – unalerted identification response	Adequate similarity	
		a.(2) Record observing a swirl on answer sheet – alerted two alternative responses		
		a.(3) Record observing a swirl on answer sheet – multiple-choice response		TV
b. If absent	b. No action necessary	b.(1) Record observing no swirl on answer sheet – alerted two alternative or multiple-choice responses		
2. Presence or absence of pits in concrete		*a.(1) Report observing pits on answer sheet – unalerted identification response	Adequate similarity	
a. If present	a. Direct worker to remove pits	a.(2) Record observing pits on answer sheet – alerted two alternative responses		
		a.(3) Record observing pits on answer sheet – multiple-choice response		TV
b. If absent	b. No action necessary	b.(1) Record observing no pits in answer sheet – alerted two alternative or multiple-choice responses		
3. Presence or absence of pockets of water		*a.(1) Report observing pockets of water on answer sheet – unalerted identification response	Adequate similarity	
a. If present	a. Direct worker to level concrete and remove pocket	a.(2) Record observing pockets of water on answer sheet – alerted two alternative responses		
		a.(3) Record observing pockets of water on answer sheet – multiple-choice response		TV
b. If absent	b. No action necessary	b.(1) Record observing no pockets of water on answer sheet – alerted two alternative responses		
4. Firmness of concrete in response to slight pressure		*a.(1) Record on answer sheet that concrete is or is not ready following each of a number of applications of pressure – alerted two alternative responses	Adequate similarity	
a. Not firm enough	a. Direct worker to wait before using float	a.(2) Record that concrete is not firm enough, proper firmness, or too firm following each of a number of applications of pressure – multiple-choice response		TV
b. Proper firmness for floating	b. Direct worker to begin floating	b.(1) Same as a.(1), a.(2)		
		b.(2)		
c. Too firm	c. Decide to omit use of float in finishing	c.(1) Same as a.(1), a.(2)		

* Preferred type of response to test item.
** Recommended presentation format for entire critical element.

(1) Multiple-choice response - This is a typical test item response format which requires the test taker to recognize the correct alternative when it is presented with two or more incorrect distractors. While this is a common test response, it is quite limited in many types of job activities. Three important characteristics of the multiple-choice responses are;

(a) The test taker must simply recognize the correct response from a limited number of alternatives; he is not doing anything other than recognizing a correct response.

(b) The test taker is typically alerted that the correct response is present among the limited number of alternatives.

(c) The test taker is responding to a small number of alternatives which are all present at the same time.

In view of these characteristics, the multiple-choice format will be appropriate when used to test a job response in which: (1) the man on the job selects the correct action from a small number of obvious possible actions, (2) the man on the job knows in advance that one of the obvious possible actions is correct, and (3) the obvious possible actions are all possible at the same point in time. The multiple-choice format is appropriately selected for example, to measure one's knowledge of what type of hammer to use to drive in a spike (assuming that on the job: (a) the right type of hammer is there to be selected, (b) the man on the job knows that one of the hammers is right, and (c) all of the hammers can be readily observed at the same time). The multiple-choice cannot be used to measure one's ability to use the hammer (e.g., note the difference between recognizing a good golf swing and doing it) or to recognize an acceptable concrete finish, unless comparator concrete slabs are always present on the job.

(2) Alerted two-alternative response - This type of response refers to true-false, go/no-go, good-bad, accept-reject, type judgments when the test taker is aware that one of the two responses is correct. The basic differences between this response and the multiple-choice response is that there are only two possible judgments as opposed to from three to five in multiple-choice items. The discussion concerning the multiple-choice response is also relevant to this type response. This format is appropriate to test inspection requirements where a given end product is accepted or rejected; but it is

not preferred to measure job performance in which a supervisor makes go/no-go decisions during some procedure, such as when to stop vibrating concrete.

(3) Unalerted identification response - This type of response varies from the multiple-choice and alerted two-alternative response in a very important respect. The test format for this type of response is one in which the test taker observes a sequence of events (for example, a construction team building the frame of a building). He is instructed to record whenever he identifies certain types of events (for example, violations of safety precautions, deviations from construction prints, improper use of tools). One way to record this answer on a structured answer sheet is to superimpose a clock on the visual test presentation, and the test taker records the time of his identifying response. This type of response is most appropriate when measuring one's ability to identify critical events as they occur in time.

(4) Unalerted decision response - This type of response is required when a test item presents a question but does not present specified alternative responses. The correct response might be anything from a number or letter symbol to a paragraph of writing. This type of response is difficult to incorporate into a standardized objective test, but with some ingenuity it can be done for specific applications. For example, if a construction drawing or a picture of a structure were included as answer sheet, some answers concerning interpretation of construction drawings could be marked on the drawing or the picture. This type of response more closely approximates the typical job situation in which one must correctly interpret something or make a decision and the correct response is not explicitly provided as one of a number of alternatives.

EXAMPLE - Column 3 of Table B-6 (Step (19)) presents possible and recommended test responses to each of the four stimulus variables. These are determined by considering both the response on the job (Column 2 of Table B-6), and the types of test responses which have been discussed in the preceding paragraphs. The recommended test response to Stimulus Variables 1, 2, and 3 in Table B-6 is an unalerted identification response, which

means that the test taker should be required to identify the presence of swirls, pits, or pockets of water if they occur as part of a test item that shows the placing and finishing of concrete. The test taker should be told at the onset of the test to note any events which require correction as they are observed, but he should not be presented a sequence of presentations and asked specifically if swirls, pits, or pockets of water are present. The reason for this is that on the job the supervisor who directs the placing and finishing of concrete must likewise detect imperfections as they occur in time; he does not have a discrete set of alerted times in which to look for a swirl in a given area or pit in a given area. Further, his response to observing the imperfection is to report his observation to the worker who corrects it. In the test situation the report is made on paper.

Note also that in the job situation, the supervisor does not respond to the absence of imperfections. As long as no imperfections are observed, work continues without input from the supervisor. Consequently, it is not recommended that a test taker be required to respond to a test item in which he reports the absence of imperfections.

An alerted two-alternative response is recommended for the fourth stimulus variable because that is the type of response on the job. The supervisor or worker intentionally applies pressure as a test of firmness, and at that instant the supervisor provides a yes-no response as to whether or not floating should begin.

(20) Compare the on-the-job versus the test response requirements to each stimulus variable and evaluate the effects of differences. The purpose of this step is to determine whether responses on the job and on a test are sufficiently similar to enable a valid test. This judgment will be in large part subjective, although greater objectivity may be possible after experience has been gained. Three situations in which response dissimilarity could have major negative effects on test validity are considered.

(1) Critic versus actor response - This refers to a situation in which the job requires a task to be done (e.g., vibrate concrete) and the test requires one to observe someone else do the job and evaluate good or bad points. For the test to be valid it is necessary to assume that recognizing mistakes is the same as not making mistakes. One should be hesitant to make this assumption - in sports

or the performing arts it is obvious that critics cannot necessarily perform although they are proficient at recognizing flaws. This type of response dissimiliarity is most likely to arise when attempting to test an action component via audio-visual simulation, and that is a major reason why Step [4] indicates performance tests should be used if only action components are involved.

(2) Level of distraction of job versus test - Some jobs require one to attend to many different things despite a variety of demands or interruptions. For example, an electronics troubleshooter may have to interrupt his troubleshooting to study circuit theory, to go to a parts manual or to go pick up tools or test equipment. The switchboard operator/receptionist must respond to many calls coming in and terminating as well as people coming and going. Any test which requires a response to some isolated task will run the danger of not providing valid results because of this type of discrepancy between the job environment and the test environment.

(3) Differences in difficulty of job and test response - This situation is similar to the preceding but not the same. In this case both the job response and the test response may have the same amount of distraction or interruption, but the job response is often more difficult. This may occur because there really are a large number of job responses (e.g., a carpenter may hammer many types of nails into many types of wood from many different positions) but a test must generalize from only one or a small number of test responses (e.g., hammering one nail into one type of wood while in one position). Hammering may be more difficult under some conditions than others, and a test which selects a less difficult condition may be of limited validity.

Another example of an unrealistically simple test response can occur if the test alerts the test taker to something that he must identify while on the job when he is not alerted. This is why Step (19) discusses alerted and unalerted responses in some detail.

After considering the differences and similarities between the job response and test response it is necessary to decide whether the responses are adequately similar to provide valid test results. This judgment will be subjective.

EXAMPLE - The judgments called for in Step (20) are listed in Column 4 of Table B-6. In deciding that each of the four preferred test responses are adequately similar to the job response the following points are considered.

(1) If the test item depicts the placing and finishing of concrete from the vantage point of a supervisor, the test taker can report his observations of defects (swirls, pits, pockets of water) as he would on the job. The test taker will be a little more alerted than he will be on the job, because the test instruction will tell him to note any observed defects. Nonetheless his response will be relatively unalerted.

(2) The job response is a "critic" response as is the test response.

〈21〉 Are the response requirements so different that you believe A/V simulation will not yield valid test results? This question appears at this time because a "YES" response ends further consideration of an A/V simulation test and virtually dictates the necessity of a performance test.

[22] YES. The answer selected shows that the test responses are sufficiently different from job responses to destroy test validity. Since response options for written tests are similar to, if not more restricted than, response options for A/V testing, a "YES" answer indicates that performance testing is necessary for valid results.

[23] NO. If this answer is selected then the use of an A/V test is quite likely.

EXAMPLE - Column 4 of Table B-6 indicates that the test responses to all four stimulus variables are adequately similar to provide valid test results. The answer to Step 〈21〉 is therefore "NO."

TEST MODE SELECTION

GENERAL

Up to this point it has been determined that an A/V simulation test will be considered if it will provide a valid test of at least one stimulus variable associated with a critical task element. At this time it is necessary to decide whether or not to use the A/V test mode. In making this decision one should

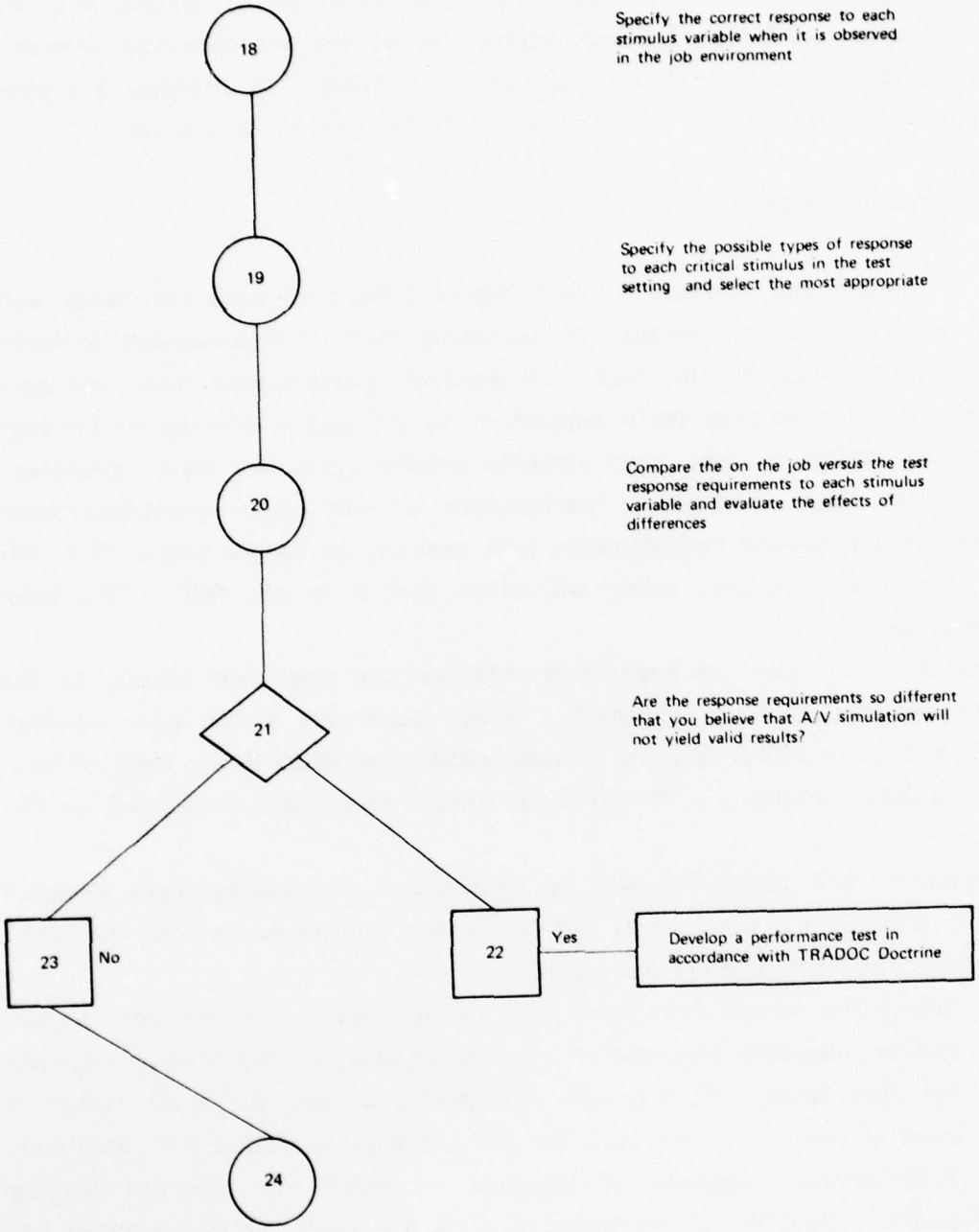


Figure B-4. Response Realism

consider factors such as cost, motivation, the feasibility of testing selected stimulus variables and not testing others, or of testing different stimulus variables by different test modes. Figure B-5 (Steps 24 through 27) provides the sequence of actions followed to select the proper test mode.

SIMULATION ALGORITHM

(24) Consider the advantages and disadvantages of each test mode and assign a test mode or test modes. In assessing cost, it is necessary to know specific characteristics of the test. In general, performance tests are most expensive. A/V or written tests supported by extensive drawing or photography will be similar in cost, and straight written tests are least expensive. Regarding test taker motivation, undesirable effects occur sometimes when test stimuli and response requirements lack realism. In simple terms, if a test does not look valid, a test taker will often feel it is not fair. This then lowers motivation.

Motivation may also be negatively affected for some test takers by the amount of reading that is required. Many Americans today have reading problems, and tests which require reading skills are difficult for them without regard to subject content. The problem reader may react negatively to the written test.

Frequently A/V simulation will be appropriate for testing some stimulus variables and critical elements but not for testing all components of the task. In such cases the following options exist:

(1) Select the single test mode that is appropriate to the most highly critical elements and stimulus variables and test all critical elements via that mode. The option still requires that the mode which is used provide a valid test for all components which are included.

(2) Test critical elements or stimulus variables via different testing modes. Usually, if performance tests are used, neither written nor A/V simulation will be needed; but on occasion it may be desirable to probe more deeply into decision or perceptual components of a task by supplementing the performance test with a written or A/V simulation test.

(3) Use the A/V simulation to test isolated stimulus variables. While some job context may be included in the simulation, responses will

be made to only the designated stimulus variables. Performance on this type of test cannot be generalized to performance of the task.

The end product of this step is a recommendation of testing mode(s) for each stimulus variable and for the critical element as a whole.

EXAMPLE - Column 5 of Table B-6 presents recommendations of TV tests for four of five stimulus variables and for the entire critical element. Note that in this case it has been decided that an adequate test of a critical element is possible even if every stimulus variable isn't tested; also that it is not recommended that an attempt be made to test, "uniformity of color of concrete" via TV. This is because it is strongly suspected that a TV presentation of this stimulus variable will be quite degraded and have an adverse effect on test taker motivation. Thus no test of that stimulus element is preferred to an inappropriate test. In considering this particular critical element, it is likely that: (1) it will be more expensive to use a performance test for this critical element alone since many concrete slabs will have to be poured to fulfill the test requirement and (2) a written test though less expensive would be much less realistic and probably less valid.

<25> Has A/V simulation been selected? This question is asked at this point because a "NO" answer eliminates the need for further concern with A/V simulation. A "YES" answer leads to further consideration in the A/V simulation test development. The answer in this step is dictated by the results of Step (24) .

[26] YES. If this answer is selected, the decision is to develop an A/V simulation test and leads to other considerations in test development. It is now necessary to proceed to Step (28) .

[27] NO. If this answer is selected, an A/V test is determined inappropriate and it is necessary to develop written and/or performance tests for the particular element.

REPRESENTATION OF JOB CONTEXT

GENERAL

Earlier evaluations have determined that A/V simulation will probably provide an adequate scoreable unit. However, the production of an A/V test

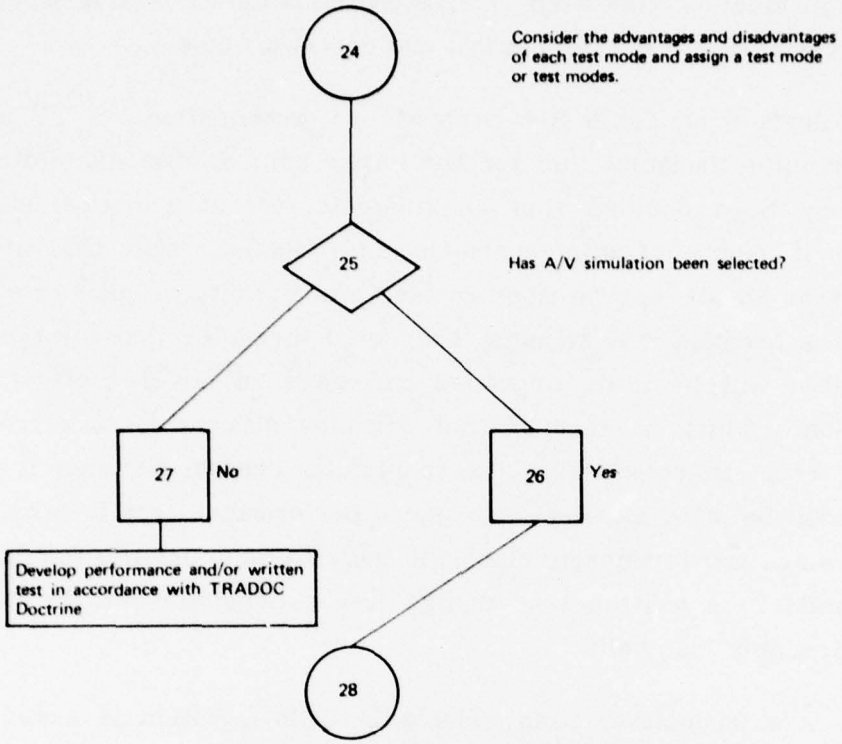


Figure B-5. Test Mode Selection

is costly and should not be used unless it will provide a reliable and valid test. To be certain it will accomplish what is desired, the task must be viewed in the job environment, the events studied and a determination made of how or if such events can adequately be presented.

Figure B-6 (Steps 28, 29, and 30) depicts a sequence of operations that will determine the adequacy of A/V simulation for the selected task.

SIMULATION ALGORITHM

(28) Study the job-functional-context in which this task is performed and identify events and stimuli that are present and that may be related to task performance. Up to this point, attention has been narrowly focused upon stimuli and responses that are immediately identified by analyzing the statement of the task and its critical elements. Now it is necessary to study the context in which the task occurs and identify other events and stimuli that influence performance of the task.

Some of the events that may occur which would influence performance in placing and finishing concrete are, for example; wind conditions, rain, or hot or cold temperatures. Therefore, such events must be considered as part of the job environment. To recognize the events, certain stimuli must be present (for example; temperature is recognized by either a scale reading on the thermometer or warmth and cold as felt by the body; wind conditions may be recognized by observing objects blowing).

In accomplishing this step at least two subject matter experts should observe a qualified person perform the task in a typical job setting. During this observation attention should be focused upon the effect that people and conditions (events) have upon task performance. The end product should be a list of contextual events and stimuli that should be considered.

EXAMPLE - The critical element, "Direct/Control Finishing Concrete Using a Wood Float", is influenced by a number of variables that have not been considered previously. These are the events and stimuli that influence the job performance. A partial list has been entered in Columns 1 and 2 of Table B-7. This list provides the basis for the following steps.

(29) Determine the importance of each contextual event to task performance. The purpose of this step is to further define the simulation requirements. Experience has shown that it is not necessary to simulate total environment for valid use of simulation in training or testing, but it is necessary to include stimuli that relate directly to the individual's performance. The same basic question presented earlier in Step (13) is repeated in this step. In this step it is necessary to review the list of contextual events which were developed in Step (28) and to determine their importance to task performance. "Considering all of the information which is provided by all of the contextual events which are typically present when this critical element is performed, how often does this one contextual event provide unique information that is essential to proper performance of the task element?"

An event will be very important if it always provides unique information which is essential to performance. It will be less important if it either duplicates information already available or if its informational value has a less direct effect upon proper performance. Each contextual event should be labeled as very important, moderately important, or not very important.

TABLE B-7. EVALUATION OF THE JOB-FUNCTIONAL-CONTEXT

Critical Element: Direct/Control Finishing Concrete Using a Wood Float

(1) Contextual Event (Step 28)	(2) Contextual Stimuli (Step 28)	(3) Criticality of Contextual Event (Step 29)	(4) Inclusion of Event in Test (Step 30)	(5) Stimuli Selected for Presentation (Step 30)
Type of concrete	Oral and written statement	Very important	Yes	Oral and written statement
Temperature	Thermometer, warmth or cold as felt	Very important	Yes	Thermometer plus oral statement
Humidity	Rain, humidity as felt	Very important	Yes	Written plus oral statement
Wind speed	Wind as felt by body, observation of items blowing	Very important	Yes	Observation of flag blowing plus oral statement
Time since placing concrete	Watch face, performance of other tasks of known duration	Very important	Yes	Watch face, oral statement

EXAMPLE - In performing this step the five contextual events which are listed in Column 1 of Table B-7 are reviewed. A judgment is made of the importance of each event in performing the critical element. In Column 3 of Table B-7, this judgment is recorded. A contextual event is rated: (1) very important if it provides unique information that is essential to proper performance, (2) moderately important if it provides information which is either unique or essential, but not both, and (3) not very important if it provides information which is neither unique nor essential to proper performance.

(30) Identify which contextual events shall be included and how they will be presented. All very important events should be represented in the test. Moderately important or not very important events may be included to the degree that the added realism is compatible with cost and time constraints. Often contextual information cannot be presented in real time or with a high

degree of stimulus realism, but providing the information in oral or written form may still enhance the validity of the test. It is also noted that many events are experienced in terms of a number of related stimuli.

EXAMPLE - In this example all of the contextual events which are identified in Step (28) are judged very important in Step (29) . Thus all are included. Since most events in real life are represented by a number of stimuli, it is desirable to provide redundant stimuli on the test. Column 5 of Table B-7 requires oral plus visual stimuli in presenting each contextual event to increase the probability that these events will be recognized.

FINAL ASSESSMENT OF PRESENTATION REALISM

GENERAL

All simulation requirements have now been specified. Before developing the test, it is advisable to check any doubts about the realism with which any of the stimulus variables can be presented via the simulation. This section specifies the steps to be taken if there are any concerns. The sequence of actions is shown in Figure B-7.

SIMULATION MODEL ALGORITHM



Are you certain that the A/V simulation will provide an adequate presentation of all important stimulus variables? To get to this step, the adequacy of A/V simulation has been judged at least probable. This question is inserted to encourage the production and evaluation of the stimulus variables whenever there is doubt about the adequacy of the A/V presentation. A high degree of certainty may exist when one has previously produced or observed A/V simulation of some stimulus variable, but when experience is absent a prototype tape will be worthwhile.

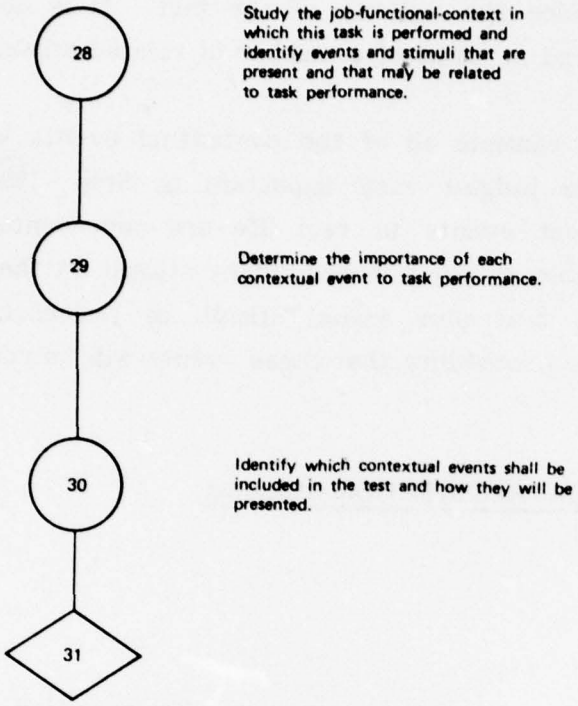


Figure B-6. Representation of Job Context

[32] NO. If this answer is selected in response to the question asked in Step <31>, it will require the development of a prototype tape that will enable a judgment of the adequacy of simulation of specific stimulus variables.

EXAMPLE - One may not be certain that TV adequately represents the effect of placing a slight pressure on concrete before and when it is ready for floating. If this is the case, development of the test is premature and a prototype tape is recommended.

[33] If the answer to this question is "YES", proceed to Step (37).

<34> Are the stimulus variables presented with adequate reality? At least two subject matter experts should be involved in answering this question. The

answer will be subjective. The main consideration in answering is whether the quality of simulation is such that a test taker will either (1) miss an item because the stimulus variable is too ambiguous or unrealistic or (2) get an item right because the simulated stimulus is too obvious. If it is suspected that the quality of simulation will either increase or decrease the probability of a correct response to the stimulus variable, simulation is not adequate.

EXAMPLE - If a prototype tape shows one applying a slight pressure to the concrete, the subject matter experts must decide whether the portrayed response of the concrete to the pressure is as clear as when looking at the actual concrete. He must ask himself the following: If it is not clear, is it bad enough to confuse the test taker? If it is clearer than what is typically observed, will this result in more test takers doing better?

[35] YES. If a "YES" answer is selected in either Step [33] or this step, it will establish the appropriateness of the intended mode of A/V simulation and it is time to begin developing the test.

[36] NO. If this answer is chosen, it rules out the intended mode of A/V simulation and directs the use of performance or written tests.

TEST DEVELOPMENT

GENERAL

The preceding steps have provided a detailed procedure for selecting and determining the adequacy of A/V simulation as a mode of testing a specific task element. Figure B-8 (Steps 37 through 54) presents the sequence of operations required to develop the test. This consists of selecting the sequence of critical elements, writing the audio-visual script, developing scoring procedures and determining the reliability and validity of the test.

SIMULATION ALGORITHM

(37) Sequence the selected critical elements in terms of test items. An A/V simulation test may include a number of items or task elements.

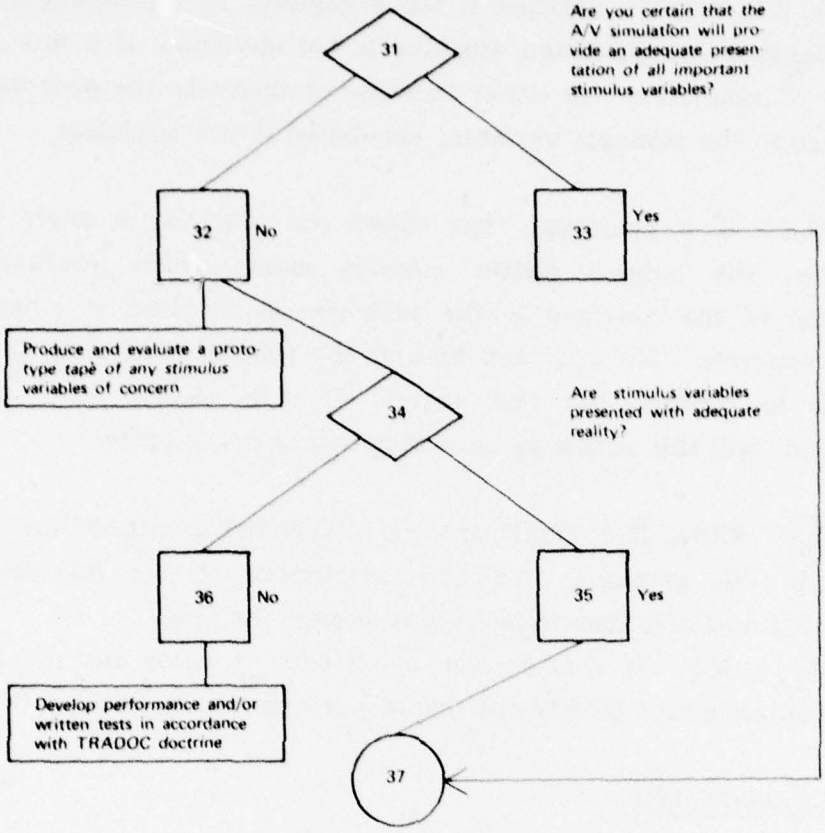


Figure B-7. Final Assessment of Presentation Realism

Ordinarily, on the job, critical elements occur in a definite order within task elements, and this order should be preserved. Sometimes the simulation test may include critical element items from two or more tasks. In this case both tasks and the critical elements should be sequenced as they occur on the job.

(38) Write audio-visual script for the simulated A/V test presentation. This is the first step in test construction. In Step (37) the critical elements to be tested were sequenced - this identified what has to be tested and when. Now it is necessary to specify the exact audio and visual content of the test. How much and what parts of the critical element will be presented in each test item and what will the test taker be told via audio about the critical element and his response? This task requires a media specialist, a subject matter expert and someone who can write the audio script in simple and unambiguous technically accurate language. Three individuals will not be

required if one person has skills in at least two of the areas. Having these resources present in the initial construction of the test item should reduce the frequency of the following types of problems.

(1) The subject matter expert states requirements for a presentation which cannot be presented via the media.

(2) The media expert fails to optimize his use of the media because he does not know the relative importance of various parts of the stimulus field.

(3) The audio script is not optimally coordinated with video.

(4) The audio script detracts from test validity because it is either technically accurate but too complex for the test taker or very clear and simple but not technically accurate.

A few general principles that should be remembered in performing this step are:

(1) Present visual test stimuli from the reference of the person performing the task. A picture of someone disassembling a rifle presents a different set of stimuli than does a picture of the rifle being disassembled from the view of the person who disassembles the rifle. A test item on assembly and disassembly of a rifle will ideally present the stimuli as they are perceived by the person doing the task rather than as they are viewed by an observer. When a critical element pertains to supervising others, the observer's reference will be appropriate.

(2) The effect of stimuli not specifically relevant to the test item must be carefully evaluated. The elimination of all apparently unrelated stimuli may destroy the job context and result in a video presentation that lacks realism. Also, sometimes a stimulus that should be trivial and insignificant can be highlighted in the simulation and become very distracting. For example, test takers may pay more attention to a soldier's cap being on crooked than to the manner in which he is performing a task. The script for the video should thus describe the video requirements in considerable detail, to include stimuli which provide context and exclude stimuli that may distract.

(3) In the job environment the soldier has a greater opportunity to obtain and verify information than he does while taking an A/V simulation test. For example, he can take a second look or ask a buddy if he sees something in the same way. The test taker will not have this freedom to verify information on his own initiative in the A/V simulation test. This emphasizes the requirement for sufficiently lengthy and clear presentations of all important stimulus variables and suggests that repeated exposure of some stimulus variables may add to test validity.

(4) The script must integrate response requirements into the A/V tests. The video need not shut off when the test taker is supposed to respond, but the script should insure that the test taker knows when and how to respond and that he will not miss other information while responding.

(5) Scoring procedures require at least two responses to each critical element. In general more than two responses are preferred.

(39) Develop criterion referenced scoring procedures in accordance with the Manual for Developing SQTs. In performing this step, it is necessary to refer to the current guidance for developing SQTs. There is one important variation to note. The guidance uses the task as the basic behavioral unit for scoring. In scoring performance on the A/V simulation test, the critical element is the basic behavioral unit. In using the procedures that are clearly spelled out in the manual, simply apply the procedures to the critical element instead of the task. For example, each critical element must have two or more test items associated with it and each critical element will be scored on a go/no-go basis.

(40) Review test script, format and scoring with five other subject matter experts. Before developing the test, it is desirable to obtain evaluative comments from qualified people who were not involved in constructing the test. They should review the audio and video script format and scoring procedures and judge whether the test is technically sound and appears capable of providing a valid measure of ability to perform the critical elements. In orienting the reviewers, it is necessary to: (1) clearly identify the purpose of the test, e.g., which critical elements are being measured, and (2) to emphasize that reviewers should focus on technical accuracy and clarity of presentation. Five subject matter experts should individually

review the proposed test materials, without further explanation or interpretation from the script writers. They should record all apparent inaccuracies or ambiguities. Following this, the writers should discuss each reviewer's comments with him. The script writers should attempt to modify the script format or scoring procedures to the satisfaction of each reviewer. Remember, if the subject matter expert made an "inappropriate" comment because he didn't understand the test item, the more naive test taker may well miss the item because it is also ambiguous for him.

It is often more difficult for script writers to participate in this review than it is to write the script for them. Script writers should keep in mind that:

(1) If the script requires your explanation now, it will probably also require explanation to some test taker. But - you will not be able to do that. It's better to be safe and modify the script now.

(2) The subject matter expert is not attacking you or your technical knowledge or your writing ability if he suggests a change. If a test comes out of this review step without any modifications, it is more likely a sign of sloppy review than of a perfectly constructed test.

(3) The time taken to revise at this point in the test development is minor compared to that required to modify the A/V simulation test because of inadequate validity or reliability. If in doubt, make changes at this stage of review.

(4) This step should end with all subject matter experts agreeing that when developed, the test should provide a technically accurate and valid measure of the critical elements.

◇41◇ Is the test judged technically sound? This question is asked explicitly because of its importance and because an objective basis for answering is provided. In practice this step occurs simultaneously with the conclusion of Step (40). Test developers will perform this step as they review comments from and discussions with each subject matter expert.

[42] YES. If this answer is selected following the preceding review and revision and the subject matter experts agree on the accuracy and clarity of the test items, it is anticipated that a test will be acceptable in terms of validity and reliability.

[43] NO. This answer may be selected if one subject matter expert contends that an item is technically inaccurate or unclear to him. A "NO"

answer is not required if the subject matter expert agrees to item accuracy and clarity but feels that the item should be presented or scored in another manner. An answer of "NO" requires a return to Step (38) .

(44) Develop A/V simulation test and response forms. The A/V simulation test is physically developed in this step. Audio and video recording and editing must be accomplished by personnel who are skilled in production techniques and procedures. A subject matter expert must be present to assure adherence to the audio and visual script and that any last minute modification will be technically acceptable. Technical flaws such as extraneous recording sounds or lighting changes, which are apparently not critical to task performance will often distract the test taker. Professional quality work on this step is highly important. It is beyond the scope of this model to go into the details of developing the A/V simulation.

(45) Review tests with five subject matter experts. In this review the subject matter expert will first be given the test, just as it is to be given to other soldiers. Extra instructions or background will not be provided. The tests will then be scored according to the prescribed scoring procedure.

Following this, the subject matter experts should be encouraged to discuss their feelings about the quality of the test.

Carefully study all errors on the test made by subject matter experts. There is a good chance that such errors indicate a flaw in the item. Check with the person who made the error and determine the cause.

Record all subjective comments. Favorable comments are nice to hear, but pay more attention to the unfavorable. Remember, the ultimate goal is a well-developed test.

◇46◇ Is the A/V presentation of acceptable quality? This step occurs simultaneously with the conclusion of Step (45) . Test developers consider this question as they analyze the test results and comments obtained from subject matter experts.

[47] YES. This answer may be given when: (1) no errors were made by subject matter experts because of inaccuracies or ambiguities in the test and (2) no more than two of the five subject matter experts agree that any specific aspect of the test is misleading, ambiguous, or distracting.

[48] NO. This answer will be selected whenever it is found that (1) a subject matter expert has missed a test item because of an ambiguity or error in the test, or (2) at least three of the five subject matter experts agree that

some aspect of the A/V simulation is misleading, ambiguous, or distracting. The rather conservative standard of three out of five is used because it is expensive to revise the test at this stage of development. An answer of "NO" requires a return to Step (44) .

◇49◇ Is the test of adequate reliability? In this step a determination is made of whether or not the test is reliable. The principle is that people who fail once should fail the second time and people who pass once should pass the second time. The method of computing reliability is contained in Chapter 7 of Developing Criterion Referenced Tests, Appendix B.

[50] YES. This answer is selected if the Ø coefficient is .50 or more. The test item has sufficient reliability for validation.

[51] NO. This answer is selected if the Ø coefficient is .49 or less. It is now necessary to return to Step (38) .

◇52◇ Is the test of adequate validity? Earlier steps established the content validity of the A/V test. Subject matter experts who possess the knowledge and skill of the MOS have judged the content of the test to be adequate. In this step it is necessary to compare individual results on the A/V test with their results on a performance test covering the same test items. This is accomplished in a manner similar to the validation procedures, currently in effect, for Phase 1 of a written test.

The validation is accomplished as follows:

(1) Validate the A/V test against a performance test for the same task. Two or more experts develop procedures for administering and scoring a performance test of the task. The procedures are refined until the experts agree perfectly in scoring. They may administer the test to each other or another individual. The scoring of the experts is the standard in subsequent steps of this type of validation.

(2) Administer both the performance test and the A/V test based on the task to groups of at least five masters and five nonmasters. The minimum acceptable standard for the performance test of the task is that 80 percent or more of the master group pass the performance test and that 20 percent or less of the nonmaster group pass the performance test. No evaluators other than the expert need observe administration of the performance test.

(3) Obtain the extent of agreement between go/no-go on the performance test and pass-fail on the A/V test. Sixty percent or more of the scores must be in agreement; that is, at least 60 percent of the soldiers pass both the performance test and the A/V test or fail both the performance test and the A/V test. A minimum of 60 percent agreement must be obtained for each A/V test item.

Assume that four A/V test items are tried out. For each of these four items, prepare a table as shown below. This is called a two by two table, where the extent of agreement is calculated between the performance test of the task and each A/V test item. In general a table is interpreted in the following manner:

PERFORMANCE TEST

A/V TEST ITEM	PASS	FAIL
PASS	CELL 1 Pass performance test and pass A/V item	CELL 2 Fail performance test and pass A/V item
FAIL	CELL 3 Pass performance test and fail A/V item	CELL 4 Fail performance test and fail A/V item

$$\text{Extent of agreement} = \frac{\text{Cell 1} + \text{Cell 4}}{\text{Cell 1} + \text{Cell 2} + \text{Cell 3} + \text{Cell 4}} \times 100$$

Now to the specific examples:

A/V ITEM 1	PERFORMANCE TEST PASS	FAIL
Pass	5	0
Fail	0	5

Extent of agreement = [(5 + 5) ÷ 10] X 100 = 100%
A/V item is satisfactory.

A/V ITEM 2	PERFORMANCE TEST PASS	FAIL
Pass	4	2
Fail	1	3

Extent of agreement = [(4 + 3) ÷ 10] X 100 = 70%
A/V test is satisfactory

A/V ITEM 3	PERFORMANCE TEST PASS	FAIL
Pass	3	2
Fail	2	3

Extent of agreement = [(3 + 3) ÷ 10] X 100 = 60%
A/V item is satisfactory

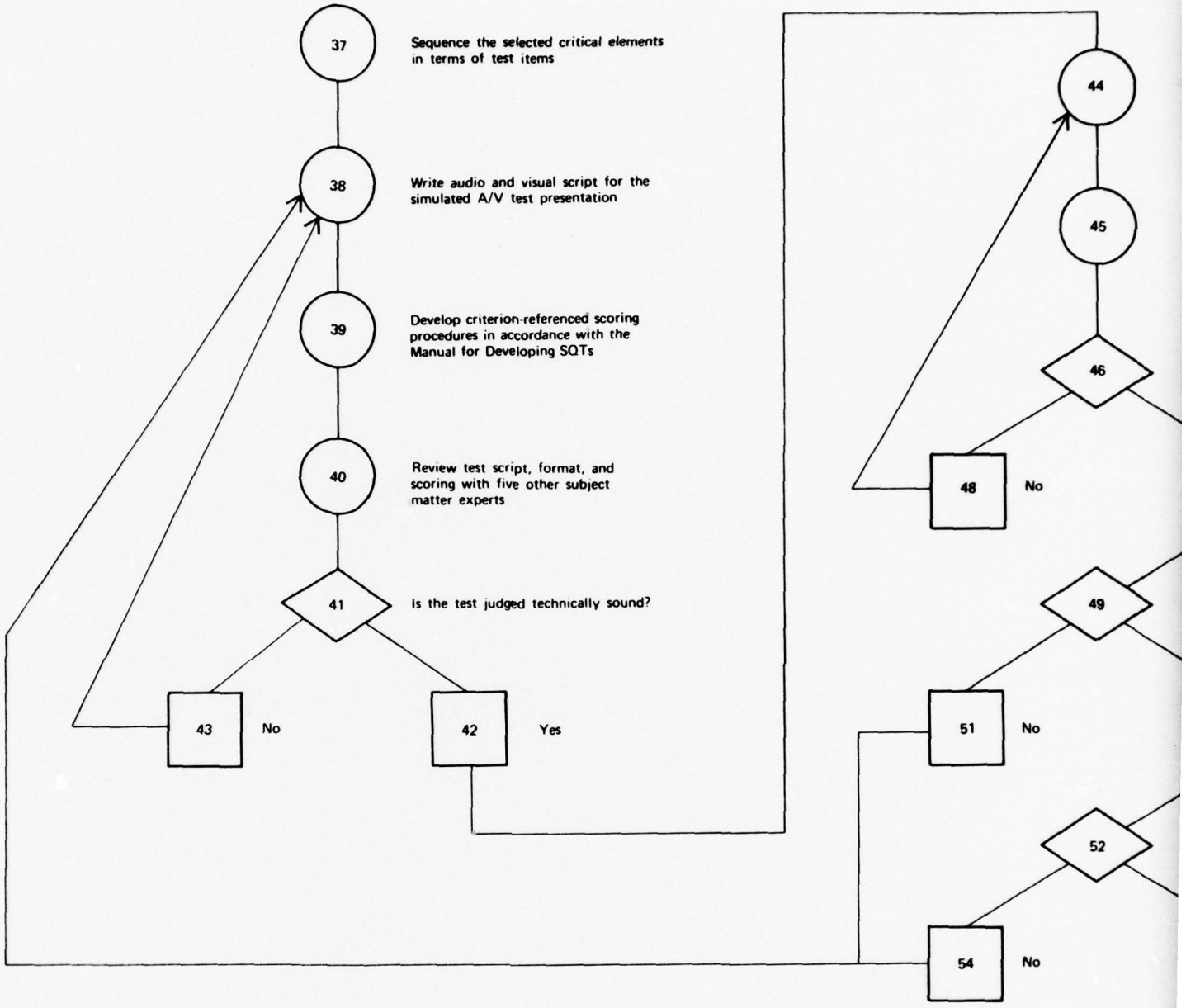
A/V ITEM 4	PERFORMANCE TEST PASS	FAIL
Pass	3	3
Fail	2	2

Extent of agreement = [(3 + 2) ÷ 10] X 100 = 50%
A/V item is unsatisfactory

Because Items 1, 2 and 3 had sufficient agreement with the performance test of the task, 60 percent or more, they were satisfactory. Since Item 4 did not meet the 60 percent criterion, it is unsatisfactory, and therefore, either requires revision or replacement by a satisfactory item. Complete this procedure of comparing A/V items based on a task to a performance test of that task for each of the A/V items.

[53] YES. If the test has adequate validity, the test development is complete.

[54] NO. This answer will be selected if the test is not determined valid and the test developer must go back to Step (38) .


37
Sequence the selected critical elements in terms of test items
38
Write audio and visual script for the simulated A/V test presentation
39
Develop criterion-referenced scoring procedures in accordance with the Manual for Developing SQTs
40
Review test script, format, and scoring with five other subject matter experts
41
Is the test judged technically sound?
43
No
42
Yes
44
45
46
48
No
49
51
No
52
54
No

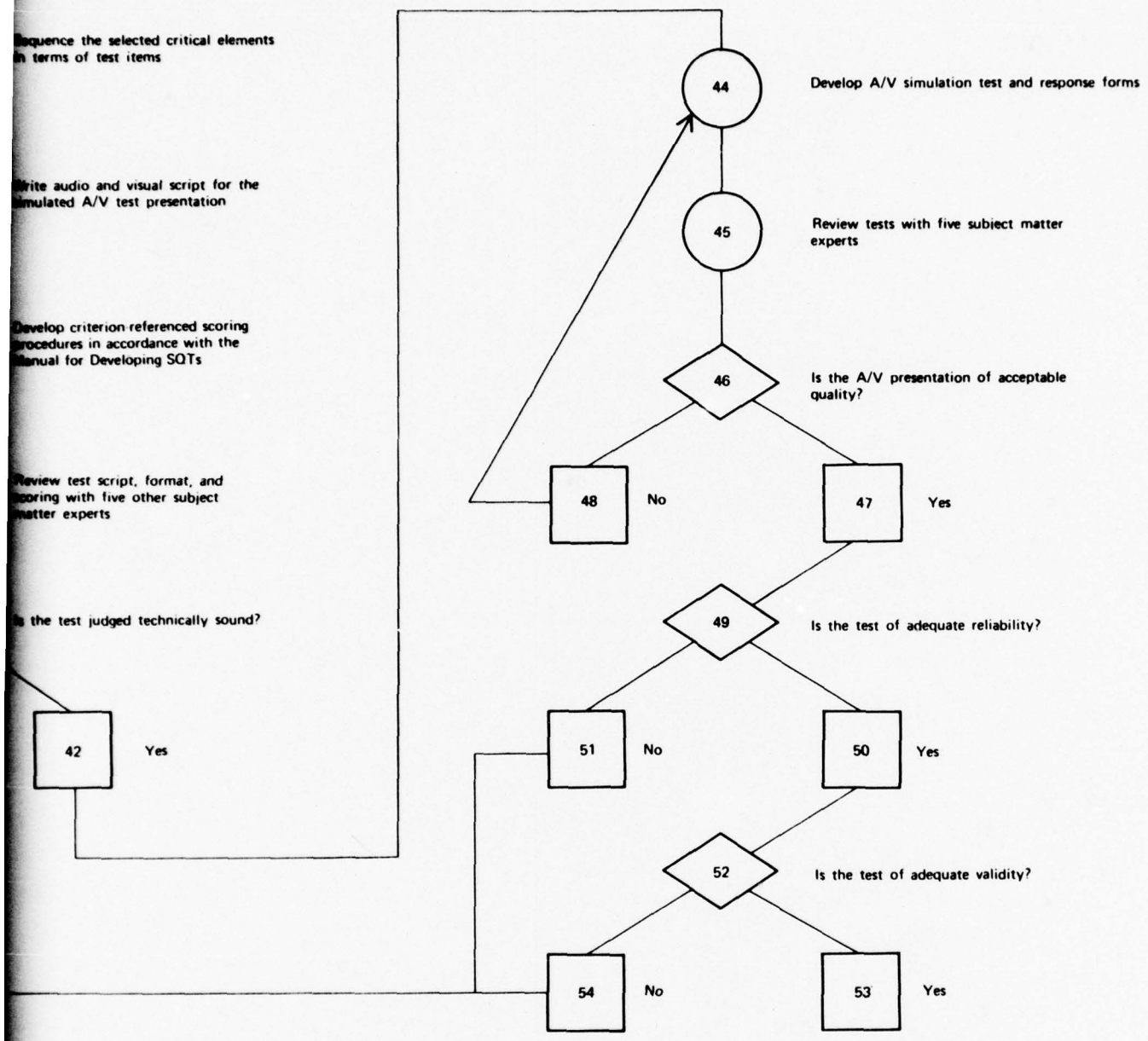


Figure B-8. Test Development

Appendix C

SIMULATION TEST AUDIO SCRIPT AND ANSWER SHEET

APPENDIX C

SIMULATION TEST AUDIO SCRIPT AND ANSWER SHEET

AUDIO SCRIPT

This is an Experimental Skill Qualification Test, S-Q-T two, for M-O-S 51B.

You have completed your training and now you are ready to be tested on the tasks "Constructing and Erecting Wall Forms" and "Placing and Finishing Concrete."

This test has been divided into three units.

Unit I: Handtool Maintenance and Materials Preparation.

Unit II: Erecting Wall Forms.

Unit III: Placing and Finishing Concrete.

The format for answering the test items will change from one unit to another, and you will be given directions telling you how to mark your answer. However, for every item on the test you will be shown a situation and a question will be stated. Then you will be given several answers from which to choose, and the question will be repeated. You will be given time to circle your answer.

You should have an answer sheet on your desk. If you do not, raise your hand and one will be given to you.

Look at the top right-hand corner of your answer sheet. You should have filled in your name, social security number and paygrade. Be sure you have completed this information before turning in your answer sheet.

Now look at the section labeled "Unit I" on your answer sheet.

Notice that in each item, there are up to four letter choices. Also there are numbers one through five to indicate "Safety Violations."

Look in the upper left-hand corner of your answer sheet. There are five possible Safety Violations:

1. Failure to ground electric tools or equipment properly.
2. Failure to wear protective gear when necessary.
3. Use of a tool in a hazardous manner.
4. Unsafe vehicle operating procedures.

and

5. Unsafe material handling or storage procedures.

If at any time during the test you see any of these five Safety Violations, circle the number that corresponds with the Safety Violation beside that item.

You will not be told when to look for Safety Violations! So be alert. Watch for them in Units one and two.

Now let's look at two sample test items.

Sample number one. You are checking plywood to see if it is usable for building a wall form. Another member of the team is cutting studs to length.

To which side of the plywood do you nail the studs?

A.

B.

To which side of the plywood do you nail the studs? Now mark your answer.

You should have circled letter "A", for sample item one on your answer sheet. Side "A" is the rough side, therefore, it is the side to which you would nail the studs. Did you notice the Safety Violations? The man using the saw was not wearing goggles.

You should have circled number two indicating failure to wear protective gear when necessary.
Now let's look at sample item number two. You are checking tools to see if they are usable.

Which tool is not usable?
A.
B.
C.
D.

Which tool is not usable? Now mark your answer.

When you noticed that the handsaw was bent, you should have circled "C" for sample item number two on your answer sheet.

Since there were no Safety Violations you should not have circled any of the numbers for that time.

Many steps are necessary to prepare for building concrete wall forms. Nine of these steps will be included in this first unit. Let's begin the test.

Item one. Jointing a saw ensures a clean cut. To joint the saw, it must be placed in a vise. For jointing, which saw is correctly gripped in the vise?
A.
B.
C.
D.

For jointing, which saw is correctly gripped in the vise? Now mark your answer.

Item two. Which man is correctly jointing the saw teeth?

A.

B.

C.

D.

Which man is correctly jointing the saw teeth? Now mark your answer.

Item three. Which man has properly gripped the handsaw in the vise and is correctly sharpening it?

A.

B.

C.

D.

Which man has properly gripped the handsaw in the vise and is correctly sharpening it? Now mark your answer.

Item four. You are building a wall form. Braces for this wall form require fourteen "sixty-seven inch" lengths of two-by-four. In order to save on materials, which stack of two-by-fours would you use?

A.

B.

Braces for this wall form require fourteen "sixty-seven inch" lengths of two-by-four. In order to save on materials, which stack of two-by-fours would you use?
Now mark your answer.

Item five. From the stack you selected, how many boards will you take to cut fourteen "sixty-seven inch" lengths?

A.

B.

C.

D

From the stack you selected, how many boards will you take to cut fourteen "sixty-seven inch" lengths? Now mark your answer.

Item six. Which man is correctly marking "twelve-inch centers" on the two-by-four?

A.

B.

C.

D.

Which man is correctly marking "twelve inch centers" on the two-by-four? Now mark your answer.

Item seven. Look at the information given on this drawing. What is the correct spreader length?

A.

B.

C.

D.

Look at the information given on this drawing. What is the correct spreader length? Now mark your answer.

Item eight. Look at these spreaders. Which one should you choose for constructing this wall form?

A.

B.

C.

D.

Which one should you choose for constructing this wall form? Now mark your answer.

Item nine. Choose the correct angle between the saw and the work.

A.

B.

C.

D.

Choose the correct angle between the saw and the work. Now mark your answer.

Item ten. Which nails are best suited for formwork?

A.

B.

C.

D.

Which nails are best suited for formwork? Now mark your answer.

Item eleven. Which example shows the correct method of hammering?

A.

B.

C.

D.

Which example shows the correct method of hammering? Now mark your answer.

This completes Unit I.

Unit II will test you on your ability to recognize the proper method for erecting the wall form.

Look at Unit II on your answer sheet. The items listed there will be shown to you in a continuous sequence.

Wall plumb

Walls level

Tie wires: holes drilled correctly
Spreader properly drilled
Wale properly installed
Spreaders in the correct position
Tie wires: twisted correctly
And, keyway positioned correctly. You are to judge whether each of the items on the list is performed correctly or incorrectly and circle the letter "C" for correct, or the letter "I" for incorrect in the appropriate space. Remember, look for Safety Violations.

Watch carefully. You will be shown this sequence only once. This time, you must circle your answer as each item is presented. Please note that each item is to be marked C-correct or I-incorrect. Let's begin.

Item twenty. Here are four methods for bracing a wall form. You will be shown the methods twice. The first time is for observation only. After the second viewing you will be given time to circle your answer. Select the best method for bracing a wall form.
A.
B.
C.
D.

Look at the braces again. Select the best method for bracing a wall form.
A.
B.
C.
D.

Now mark your answer.

This completes Unit II.

Unit III will test you on your ability to recognize the correct placing and finishing of concrete.

Look at Unit III on your answer sheet. This Unit will be completed in the same manner as in Unit I. Circle the letter you choose as the correct answer for each item. There are no Safety Violations. Let's begin.

Item twenty-one. Concrete will be poured into a form forty-eight inches high. At what level should you begin to vibrate the concrete?

A. 6 inches.
B. 18 inches.
C. 36 inches.
D. 48 inches.

At what level should you begin to vibrate the concrete? Now mark your answer.

Item twenty-two. Which is the correct way to store a wheelbarrow?

A.
B.
C.
D.

Which is the correct way to store a wheelbarrow? Now mark your answer.

Item twenty-three. Choose the proper screeding technique. You will see each choice only once.

A.
B.
C.
D.

Choose the proper screeding technique. Now mark your answer.

Item twenty-four. Is this concrete ready to be floated? You will see each choice only once.

A. Yes.
B. No.
C. The information is insufficient.

Is this concrete ready to be floated? Now mark your answer.

Item twenty-five. Which is the proper method for floating concrete? You will see each choice only once.

A.

B.

C.

D.

Which is the proper method for floating concrete? Now mark your answer.

Item twenty-six. Which slab has been properly floated? You will see each choice only once.

A.

B.

C.

D.

Which slab has been properly floated? Now mark your answer.

Item twenty-seven. Look at this concrete. Is it ready for first troweling?

A. Yes.

B. No.

C. The information is insufficient.

Is it ready for first troweling? Now mark your answer.

This completes the experimental skill qualification test for M-O-S 51B. Please stop writing and await further instructions. Thank you.

SAFETY VIOLATIONS:

1. Failure to ground electric tools or equipment properly.
2. Failure to wear protective gear when necessary.
3. Use of a tool in a hazardous manner.
4. Unsafe vehicle operating procedures.
5. Unsafe material handling or storage procedures.

NAME _____

SSAN _____

PAYGRADE _____

ANSWER SHEET
MOS 51-B

DIRECTIONS: Mark your answer by circling the correct letter and/or number.
WATCH FOR SAFETY VIOLATIONS

UNIT I

Sample: Safety Violation

1. A B C D 1 2 3 4 5
2. A B C D 1 2 3 4 5

1. A B C D 1 2 3 4 5
2. A B C D 1 2 3 4 5
3. A B C D 1 2 3 4 5
4. A B C D 1 2 3 4 5
5. A B C D 1 2 3 4 5
6. A B C D 1 2 3 4 5
7. A B C D 1 2 3 4 5
8. A B C D 1 2 3 4 5
9. A B C D 1 2 3 4 5
10. A B C D 1 2 3 4 5
11. A B C D 1 2 3 4 5

UNIT II

		Correct	Incorrect	Safety Violation
12.	Wall plumb	C	I	1 2 3 4 5
13.	Walls level	C	I	1 2 3 4 5
14.	Tie wires: holes drilled correctly	C	I	1 2 3 4 5
15.	Spreader properly drilled	C	I	1 2 3 4 5
16.	Waiers properly installed	C	I	1 2 3 4 5
17.	Spreaders: correct position	C	I	1 2 3 4 5
18.	Tie wires: twisted correctly	C	I	1 2 3 4 5
19.	Keyway positioned correctly	C	I	1 2 3 4 5
20.	Wall braces	A B C D		1 2 3 4 5

DIRECTIONS: Mark your answer by circling the correct letter and writing in the correct rating.
NO SAFETY VIOLATIONS

UNIT III

21. A B C D
22. A B C D
23. A B C D
24. A B C D
25. A B C D
26. A B C D
27. A B C D

Appendix D

WRITTEN TEST

APPENDIX D

WRITTEN TEST

DIRECTIONS: Choose one answer to each item and circle the appropriate letter.

NAME _____

SSAN _____

PAYGRADE _____

Hand Tool Maintenance and Material Preparation

1. For jointing, the saw is placed in a vise. Which is the correct position for gripping the saw?
   a. Place the saw in the vise with the teeth about 2 inches above the vise jaws.
   b. Place the saw in the vise so the gullets of the teeth are about 1/4 inch above the edge of the vise jaws.
   c. Place the saw in the vise so that the blade sits at an angle - the heel should be higher than the toe.
   d. Place the saw in the vise so that the jaws of the vise grip the bottom of the saw blade.
2. Which is the correct method for jointing the saw teeth?
   a. Place a triangular file in the jointer and move it lightly over the saw teeth from heel to toe, without rocking the file.
   b. Place a mill file in the jointer and move it lightly over the saw teeth from heel to toe, without rocking the file.
   c. Place a mill file in the jointer and while applying pressure, move it over the saw teeth from heel to toe, without rocking the file.
   d. Place a mill file in the jointer and lightly rock it back and forth over the saw teeth, moving from heel to toe.

PRECEDING PAGE BLANK

3. What is the correct procedure for sharpening a crosscut handsaw?
   a. Hold the file at a right angle to the saw blade and begin at the heel and work toward the toe.
   b. Hold the file at a right angle to the saw blade and begin at the midpoint of the saw and work toward the heel; then go back to the midpoint of the saw and work toward the toe.
   c. Hold the file at a 45^{o} to 60^{o} angle to the saw blade and begin at the heel and work toward the toe.
   d. Hold the file at a right angle to the saw blade and begin at the toe and work toward the heel.
4. Braces for a wall form require 2x4's cut to 67" lengths. In order to save materials, from which stack of 2x4's would you take your lumber?
   a. 12' - 2x4's.
   b. 8' - 2x4's.
5. From the stack you selected in the above question, how many boards will it take to cut 14 - 67" lengths?
   a. 4 boards.
   b. 7 boards.
   c. 12 boards.
   d. 14 boards.
6. How would you measure and mark the toe plate for studs which are to be placed 12" on center?
   a. Measure 12 3/4", draw a line, and mark an "X" on the left side of the line.
   b. Measure 12", draw a line, and mark an "X" on the left side of the line.
   c. Measure 11", draw a line, and mark an "X" on the left side of the line.
   d. Measure 12 3/4", draw a line, and mark an "X" on the right side of the line.

7. Using the information given in Figure D-1, what is the correct spreader length?
   a. 6".
   b. 8".
   c. 10".
   d. 12".
8. Referring again to Figure D-1, select the correct spreaders for constructing the wall form:
   a. 10" - 1x2 with holes drilled dead center.
   b. 10" - 1x2 with holes drilled off center.
   c. 8" - 1x2 (squared off) with no holes drilled.
   d. 8" - 1x2 (squared off) with holes drilled off center.
9. What is the best angle between the saw and the work?
   a. 10^{o}.
   b. 45^{o}.
   c. 75^{o}.
   d. 90^{o}.
10. Which nails are best suited for fabricating concrete forms?
   a. 8d common.
   b. 16d common.
   c. Finishing nails.
   d. Double headed nails.
11. To provide the greatest holding power, nails should be driven:
   a. Straight and parallel to each other.
   b. At an angle slightly toward each other.

Erecting Wall Forms

12. If 1x2 spreaders are used and each has a 1/4" hole drilled about 1 1/2" from the end, they have been:
   a. Drilled properly.
   b. Drilled improperly.

13. The 2x4 wale is installed so that the broad side faces the studding and the end extends about 6" beyond the panels. The wale has been:
    a. Properly installed.
    b. Improperly installed.
14. Holes for the tie wire are drilled so that there is one hole on either side of each stud requiring ties. Both holes are drilled above the wale. The holes for the tie wires have been:
    a. Properly drilled.
    b. Improperly drilled.
15. The tie wires are twisted from the center, using 16d nails, until all are of approximately uniform tension. There is a small loop left in the center of each tie wire and the spreaders resist movement. The tire wires have been:
    a. Properly twisted.
    b. Improperly twisted.

Placing and Finishing Concrete

16. Concrete will be poured into a 48" high form. At what level should you begin to vibrate the concrete?
    a. 6".
    b. 18".
    c. 36".
    d. 48".
17. Which is the correct position for storing the wheelbarrow?
    a. Lying upside-down on the ground.
    b. Lying on its side on the ground.
    c. Sitting upright on the ground.
    d. Tilted up against the side of a building with the wheels off the ground.

18. Which is the correct technique?
    a. Two men on either side of the slab should tilt the screed at an angle and scrape the excess aggregate from the slab.
    b. Two men on either side of the slab should position the screed flat on the slab, and beginning in the center of the slab, use a sawing motion.
    c. Two men on either side of the slab should position the screed flat on the slab, and beginning at one end, should use a sawing motion.
    d. Two men on either side of the slab should position the screed flat on the slab, and beginning at one end, should use a sawing motion. They should screed a short distance, fill in the depressions, and screed again until there are no depressions.
19. Which is the correct method for floating concrete?
    a. Hold the float in a position so that the tip can dig into the concrete. Use long strokes to smooth the concrete.
    b. With the side edge of the float, use long strokes to smooth the concrete.
    c. With the side edge of the float, use short brisk strokes to smooth the concrete.
20. A slab is ready to be floated when:
    a. The water sheen has disappeared and after stepping in and out of the mixture only a slight imprint is left.
    b. Water is standing on top of the mixture and after stepping in and out of the mixture no imprint is left.
21. When is the concrete ready for first troweling?
    a. As soon as it is set.
    b. Immediately after the concrete is floated.
    c. After the moisture sheen has disappeared from the surface.
    d. While the concrete is still fresh enough to work the water up to the surface.

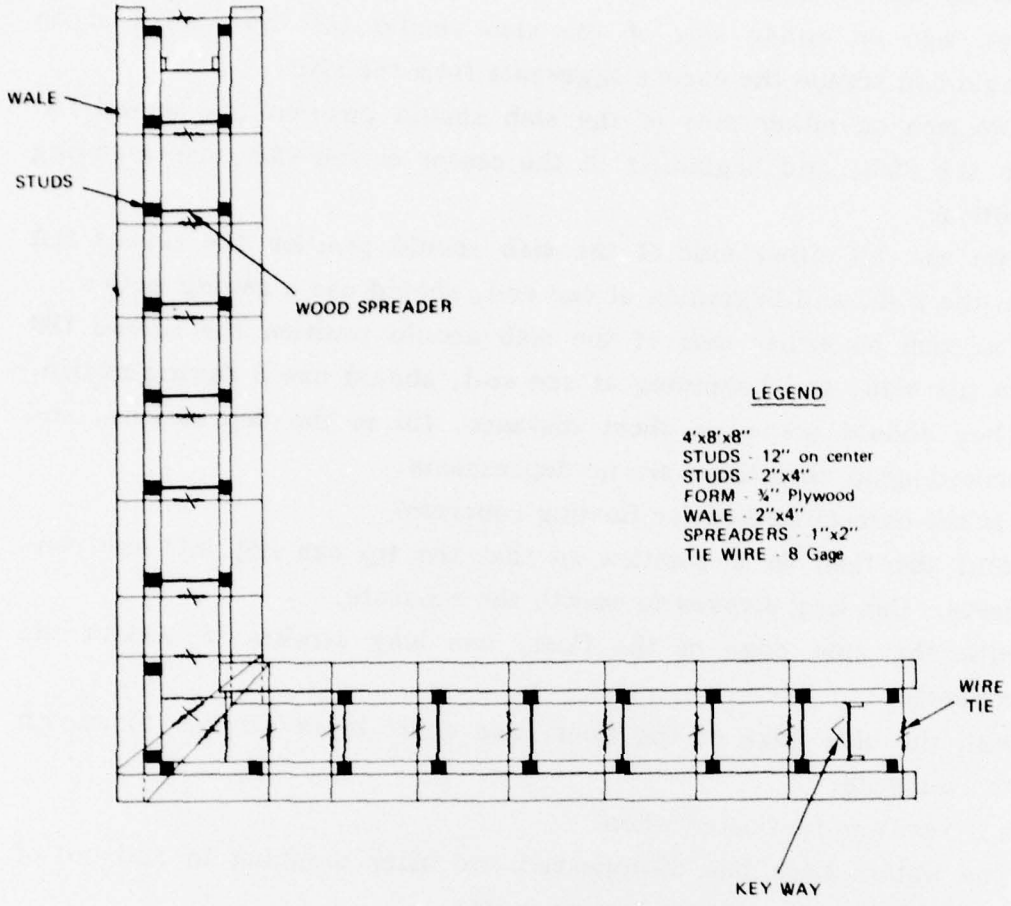


Figure D-1. Wall Form

Appendix E

QUESTIONNAIRE

APPENDIX E

QUESTIONNAIRE

The purpose of this questionnaire is to obtain your personal reactions to the television and written tests. Check the appropriate categories and fill in the information at the end of the questionnaire. Please be candid - this questionnaire is anonymous.

1. To what extent was either test a fair measure of your ability to perform in your MOS?

Television Test	Written Test
_____ Extremely fair	_____ Extremely fair
_____ Very fair	_____ Very fair
_____ Somewhat fair	_____ Somewhat fair
_____ Not fair	_____ Not fair

2. How interesting was either test?

Television Test	Written Test
_____ Extremely interesting	_____ Extremely interesting
_____ Very interesting	_____ Very interesting
_____ Somewhat interesting	_____ Somewhat interesting
_____ Not interesting	_____ Not interesting

3. How difficult was either test?

Television Test	Written Test
_____ Extremely difficult	_____ Extremely difficult
_____ Very difficult	_____ Very difficult
_____ Somewhat difficult	_____ Somewhat difficult
_____ Not difficult	_____ Not difficult

4. Overall, to what extent were the visuals (pictures, graphics, titles, etc.) clear in the television test?

_____ Extremely clear
_____ Very clear
_____ Somewhat clear
_____ Not clear

PRECEDING PAGE BLANK

5. In the case of the television test, to what extent was the narration easy to understand?

_____ Extremely easy to understand
_____ Very easy to understand
_____ Somewhat easy to understand
_____ Not easy to understand

6. In the case of the television test, to what extent was the answer sheet easy to use?

_____ Extremely easy to use
_____ Very easy to use
_____ Somewhat easy to use
_____ Not easy to use

7. Overall, did you have enough time to answer the questions to the television test?

_____ More than enough time
_____ Enough time
_____ Barely enough time
_____ Not enough time

8. What is your feeling about the overall pace (rate of presentation) of the television test?

_____ The pace was much too slow
_____ The pace somewhat too slow
_____ The pace was somewhat too fast
_____ The pace was much too fast

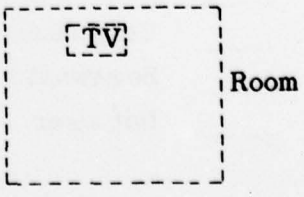
9. What is your feeling about the overall selection of items (situations) for the television test?

_____ The items were extremely well chosen
_____ The items were very well chosen
_____ The items were fairly well chosen
_____ The items were poorly chosen

10. From where you were sitting, how well were you able to see the television screen?

_____ Extremely well
_____ Very well
_____ Fairly well
_____ Not well

(Please show your seating position in the diagram on the right)



11. Can you recall any specific items (situations) in the television test that were confusing? If so, describe the item(s) in a few words. _____

_____

_____

_____

12. Do you have any additional comments on the television test? _____

_____

_____

_____

13. Can you recall any specific items in the written test that were confusing? If so, describe the item(s) in a few words. _____

_____

_____

_____

14. Do you have any additional comments on the written test? _____

_____

_____

_____