LEVEL II

12

ARI TECHNICAL REPORT
P-78-1

GUIDELINES FOR THE DEVELOPERS OF VIDEOTAPE SIMULATION PERFORMANCE TESTS

by

Alvin J. Abrams and Anthony W. Kendall
DATA-DESIGN LABORATORIES
7925 Center Avenue
Cucamonga, California 91730

May 1978

Contract DAHC 19-76-C-0032

Milton H. Maier, Technical Monitor
Individual Training and Skill Evaluation Technical Area, ARI

Prepared for

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

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.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

REPORT NUMBER: P-78-1 (18) ARI	2. GOVT ACCESSION NO.
1. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): GUIDELINES FOR THE DEVELOPERS OF VIDEOTAPE SIMULATION PERFORMANCE TESTS, (6)	5. TYPE OF REPORT & PERIOD COVERED: (7) Final rept,
6. PERFORMING ORG. REPORT NUMBER: (14) DDL - TSD-78-2002	
7. AUTHOR(S): (10) Alvin J. Abrams and Anthony W. Kendall	8. CONTRACT OR GRANT NUMBER(S): (15) DAHC 19-76-C-0032
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Data-Design Laboratories, 7925 Center Avenue, Cucamonga, California 91730	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: (16) 2Q163731A770
11. CONTROLLING OFFICE NAME AND ADDRESS: TRADOC, Army Training Support Center, Fort Eustis, Virginia (11)	12. REPORT DATE: May 1978
	13. NUMBER OF PAGES: 70
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): US Army Research Institute for the Behavioral and Social Sciences, 5001 Eisenhower Avenue, Alexandria, Virginia 22333 (12) 72p.	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

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

18. SUPPLEMENTARY NOTES

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

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)

This handbook is designed to be used by Army test developers who are considering audio-visual alternatives to the hands-on, written, and performance certification components of the Skill Qualification Test (SQT). The handbook is organized to show how to use the audio-visual medium for simulated performance tests; it covers videotapes, slide/sound combinations, and written tests with accompanying audio recordings

The handbook provides guidelines for the test developer to (1) rank

20. order critical tasks proposed for the SQT, (2) allocate the tasks to the most appropriate test medium, (3) formulate the test items to take maximum advantage of the medium selected, and (4) objectively evaluate the test.

A companion volume, ARI Technical Report TR-78-A13, details the development and evaluation of a prototype videotape simulated performance test.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (Continued)

LIST OF TABLES

SECTION I

INTRODUCTION

PURPOSE

The purpose of this handbook is to provide guidelines for the test developer to:

(1) Prioritize, or rank order, the list of critical tasks proposed for the Skill Qualification Test (SQT) in order to insure that the critical tasks determined to be appropriate for testing are considered first.

(2) Allocate the tasks to the most appropriate test delivery system through a structured, logical analysis of the behavioral elements within each task.

(3) Formulate the test items in a manner that takes maximum advantage of the test media selected.

(4) Objectively evaluate the test, within its developmental stage and as a final product.

SCOPE

The guidance contained in this handbook is designed to be used by the test developers when they are considering alternatives to the hands-on, written and performance certification components of the SQT. Because this handbook is intended to add guidance in areas not addressed by previous test development documentation, it will not overlap in detail, those areas already covered. Detailed guidance on SQT development has already been published by the Individual Training Evaluation Directorate, Ft. Eustis, VA, Handbook for the Development of Skill Qualification Tests. A detailed model for job task analysis is found in Interservice Procedures for Instructional Systems Development and further guidance is found in a wealth of civilian publications; such as Principles of Instructional Design, Gagne and Briggs, published by Holt, Rinehart, and Winston, New York.

CONTENT

The handbook is organized to show the application of the audio-visual test delivery system through the use of simulated performance tests.

Delivery Systems

The delivery systems, or media, covered in the handbook include slide/sound combinations, written tests with accompanying audio recordings and videotapes with primary emphasis on the latter method. The guidance is quite detailed and it is assumed that the reader has prior knowledge of the test development process and is conversant with all the terms common to the process; however, it is not assumed that the reader is familiar with the various media to be considered. Therefore, discussions of the various media will be largely nontechnical in nature, and will address:

- The resources necessary to produce the various simulated tests.
- Capabilities and limitations of the different media.
- Evaluative criteria for both the developmental and final product stages of the simulated tests.

Simulated Performance Test Development Procedures

(1) These procedures, in the form of algorithms, lead the reader step-by-step through the process of developing and constructing simulated performance tests. The first set of procedures offers a means of rank ordering a list of critical tasks by assessing the commonality and criticality of their component elements or performance steps.

(2) The simulation procedures provide a means for analyzing the task elements to determine the most appropriate media by which performance of the task should be measured. They give the developer an alternative (in the form of simulation) when it is found that a task is, for one reason or another, a poor candidate for any of the three standard components, i.e., written, hands-on, or performance certification.

(3) The handbook contains operational examples from a previously developed simulated performance test constructed for MOSs 51A (Utility Workers) and 51B (Carpenters).

Assumptions: It should be noted that the procedures were developed under the assumption that a thorough and complete job task analysis was made on the MOS being considered. It is further assumed that this analysis was accomplished in a standardized and systematic manner, thus insuring a uniform approach to each task when the various performance steps were developed.

SECTION II

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.

IDENTIFICATION OF COMMON BEHAVIORS

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 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 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 PROCEDURES

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, task criticality must be assessed. If the task list is taken from a soldier's manual, it can be assumed that all tasks are critical. If it is 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, 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, Alexandria, VA., 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 not to redefine task statements. By this time you should be able to consider the task statement as valid. 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 1). The selection of behavior designators is accomplished through a search of the JTSS (Figure 3), 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 Figure 3. 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) Plot designators. In this step, you 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 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. Refer to Figure 3 for example.

Notice that although the task in Figure 3 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, all critical tasks have been accounted for, sum the behavior designators horizontally across each task. This step establishes 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 shown in the lower right-hand corner of Figure 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) Circle common designators. 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 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 calibre 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 task extracted from the soldier's manual for MOS & skill level.

(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 piles 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.

TASKS					
Install roof trusses	Place & finish concrete	Install anchor bolts in concrete	C[illegible]uct joints in [illegible]oncrete	Prepare timber piles for driving	Frequency behavior designator appears in the critical tasks. See Notes 1 & 2
					1
		(X)		(X)	(5)
(X)					(5)
				(X)	(4)
X					2
			(X)	(X)	(6)
					1
					1
(X)					(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	2	3	4	55

(17) – 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.

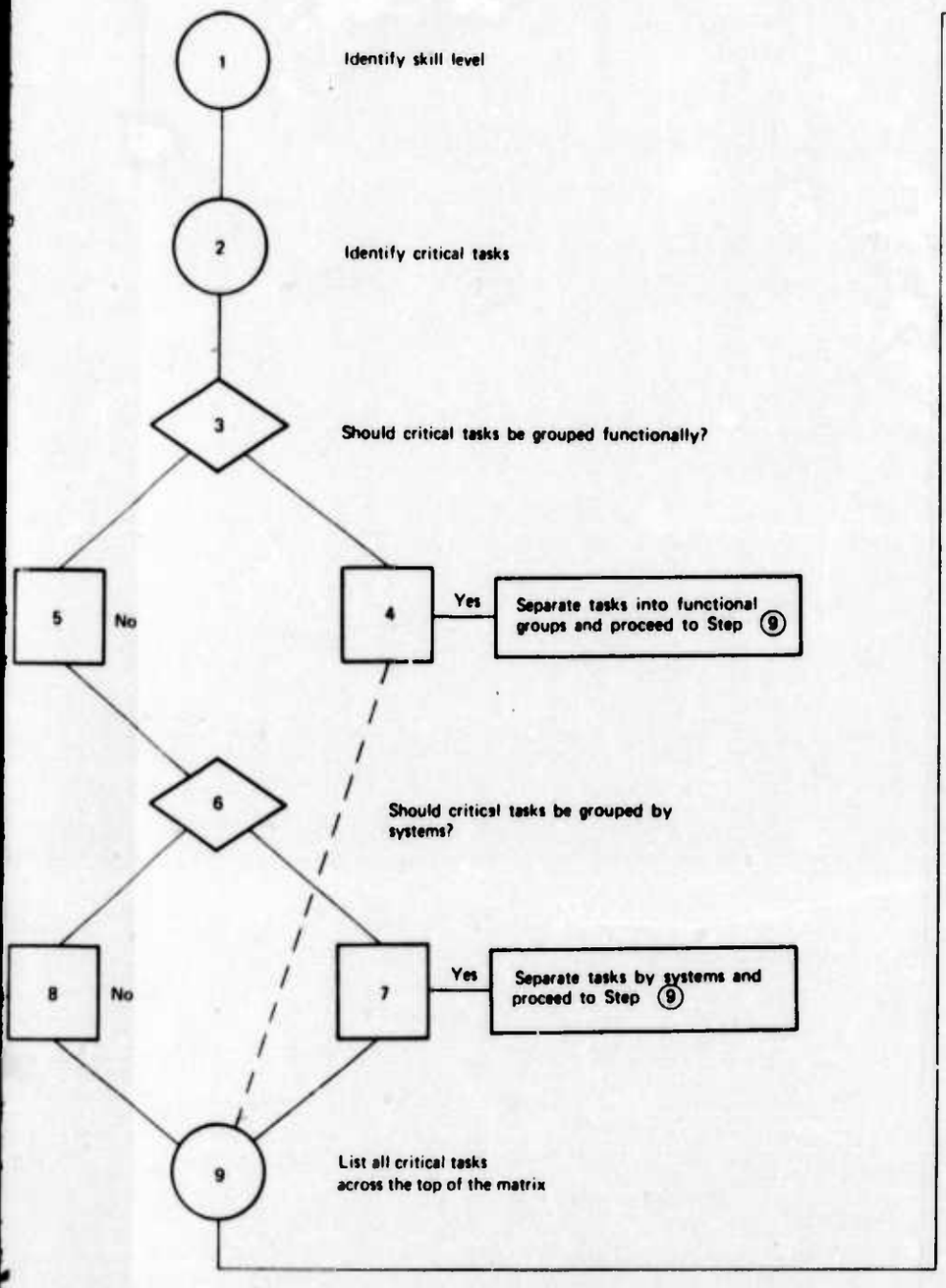
2 Numbers circled e.g., (10) correspond to the appropriate step in the algorithm & are explained in detail in the text of the handbook.

Total number of behavioral designators.

+ 23 ÷ 2.4

Average number of tasks in which a behavior designator appears.

Figure 1. Task Selection Matrix



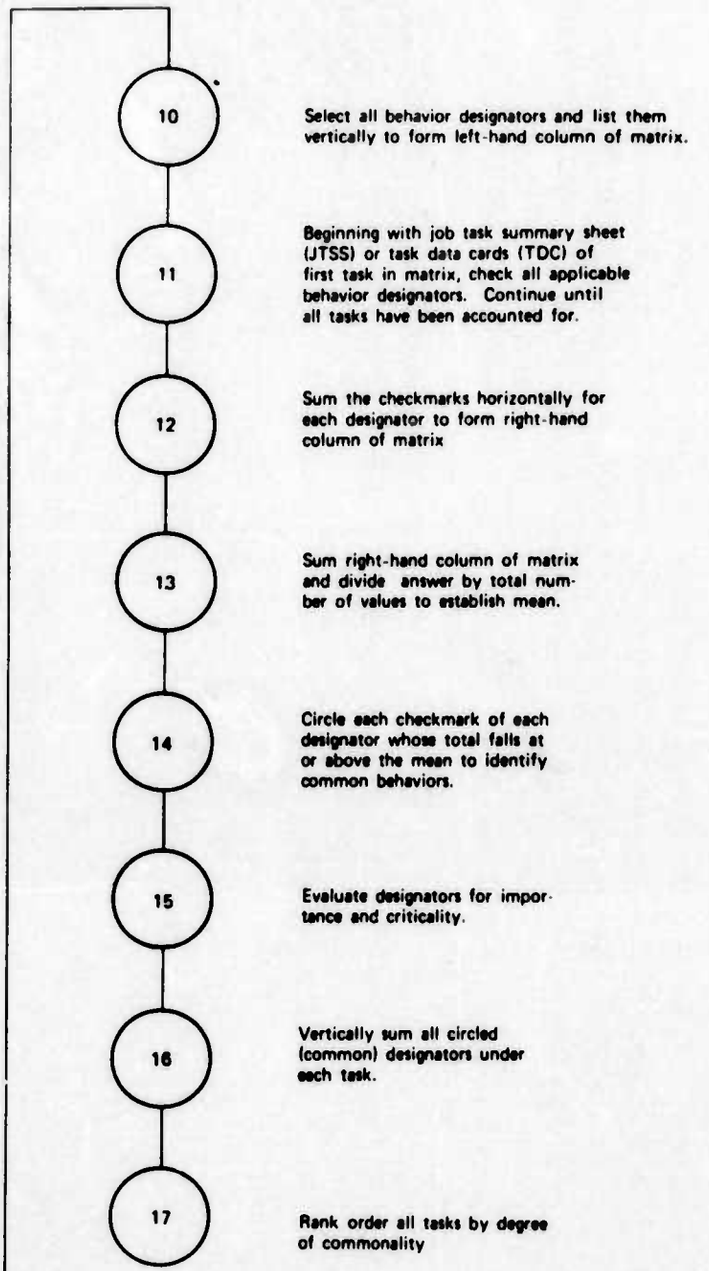
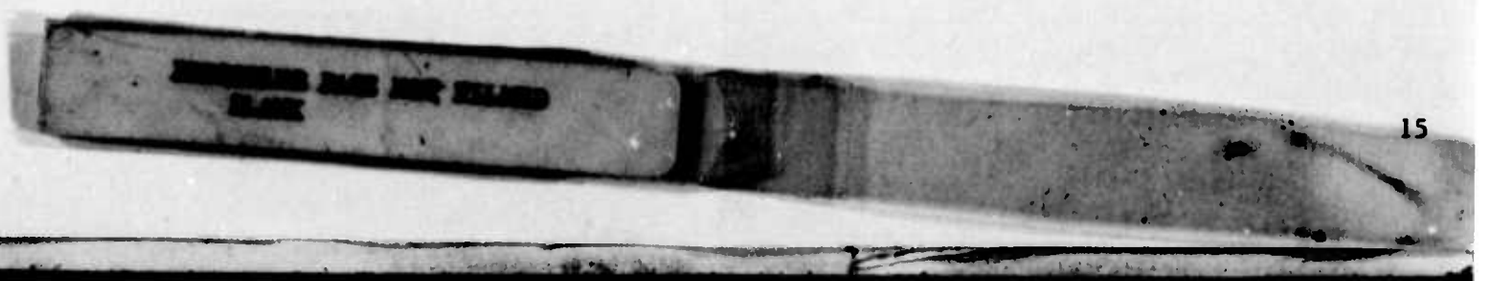


Figure 2. Task Selection Procedures



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.

Figure 3. Job Task Summary Sheet (Abbreviated)

SECTION III

SIMULATION TEST PROCEDURES

INTRODUCTION

The tasks to be used in developing the SQT have been ranked, and the following procedure will aid you in developing an audio-visual simulated performance test. You should apply these procedures to each task in order of task rank.

The procedures are presented as an algorithm (Figure 4) and are 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, ▭ . You will find the procedures are easy to follow. There are 42 steps and an explanation is provided for each step.

PRELIMINARY TEST MODE SELECTION

Background

Your application of the Task Selection Procedures provided a rank ordered list of tasks to be considered in developing the SQT. These procedures aid in determining whether a task is suitable for testing by A/V simulation.

Figure 5 depicts steps 1 through 5 and shows the sequence of operations involved in making a preliminary selection of test mode.

Simulation Algorithm

(1) Identify the critical elements. 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

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.

Figure 3. Job Task Summary Sheet (Abbreviated)

SECTION III

SIMULATION TEST PROCEDURES

INTRODUCTION

The tasks to be used in developing the SQT have been ranked, and the following procedure will aid you in developing an audio-visual simulated performance test. You should apply these procedures to each task in order of task rank.

The procedures are presented as an algorithm (Figure 4) and are 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, ▭ . You will find the procedures are easy to follow. There are 42 steps and an explanation is provided for each step.

PRELIMINARY TEST MODE SELECTION

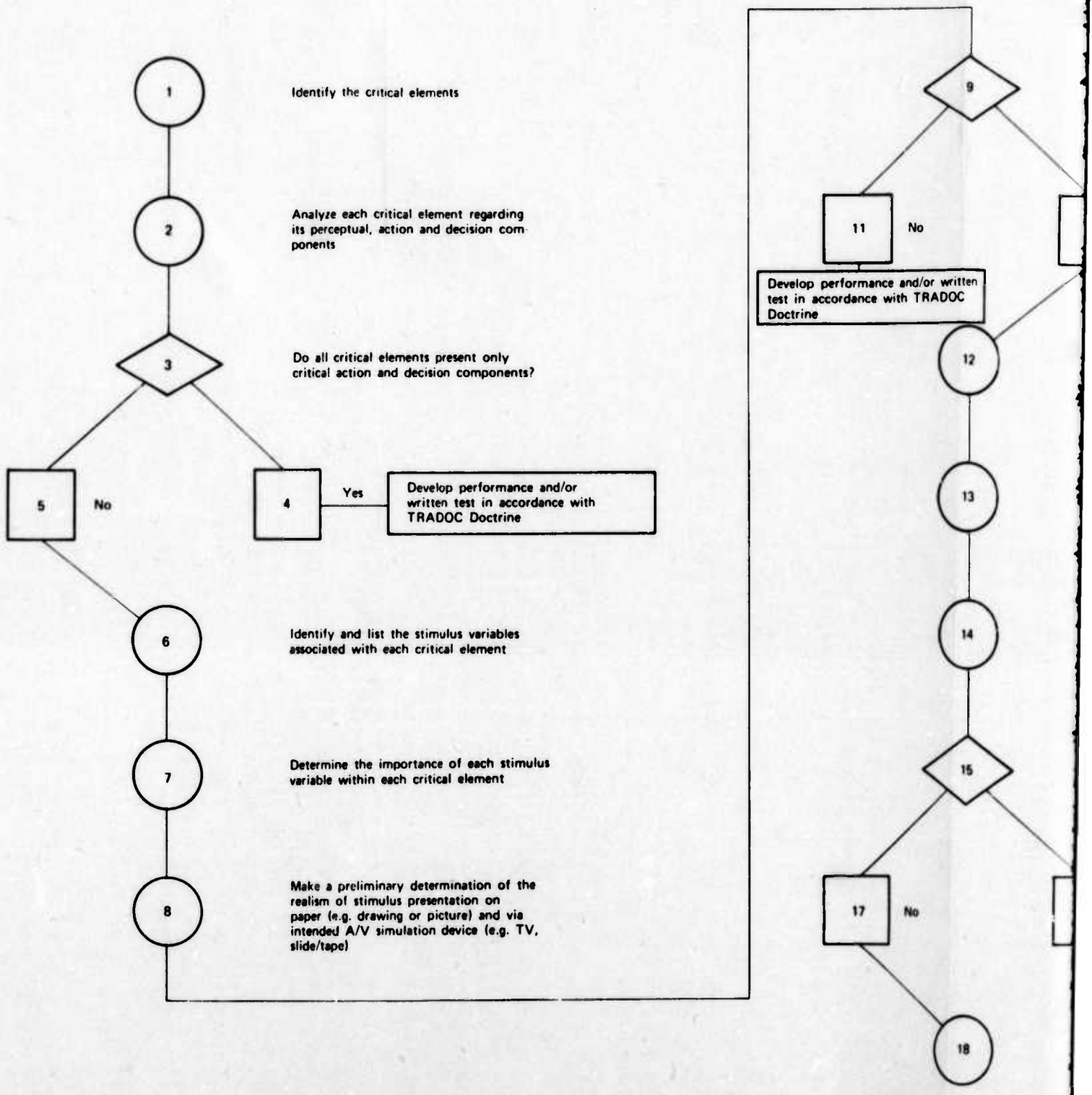
Background

Your application of the Task Selection Procedures provided a rank ordered list of tasks to be considered in developing the SQT. These procedures aid in determining whether a task is suitable for testing by A/V simulation.

Figure 5 depicts steps 1 through 5 and shows the sequence of operations involved in making a preliminary selection of test mode.

Simulation Algorithm

(1) Identify the critical elements. 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


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 and decision components?
5
No
4
Yes
Develop performance and/or written test in accordance with TRADOC Doctrine
6
Identify and list the stimulus variables associated with each critical element
7
Determine the importance of each stimulus variable within each critical element
8
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)
9
11
No
Develop performance and/or written test in accordance with TRADOC Doctrine
12
13
14
15
17
No
18

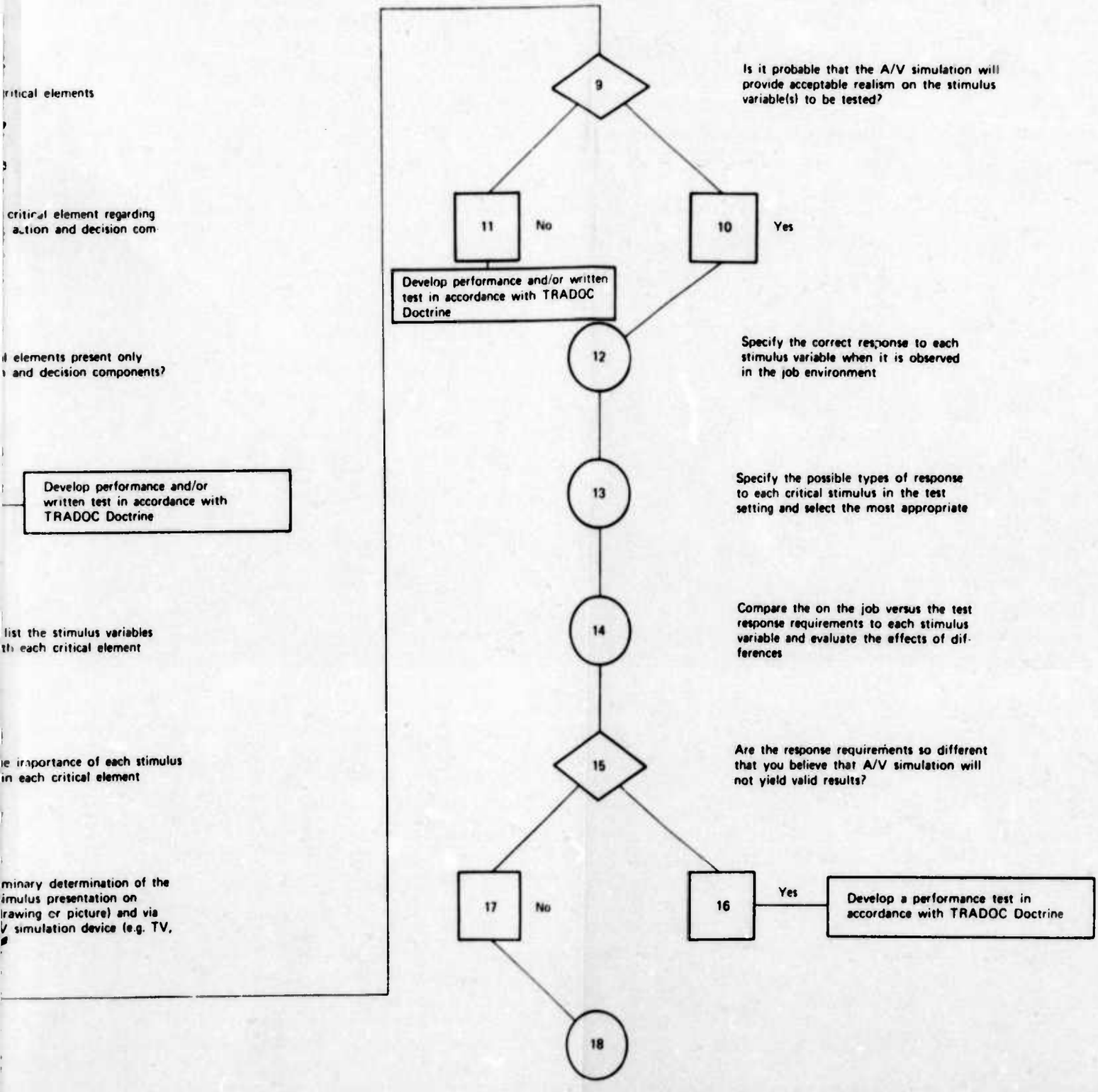
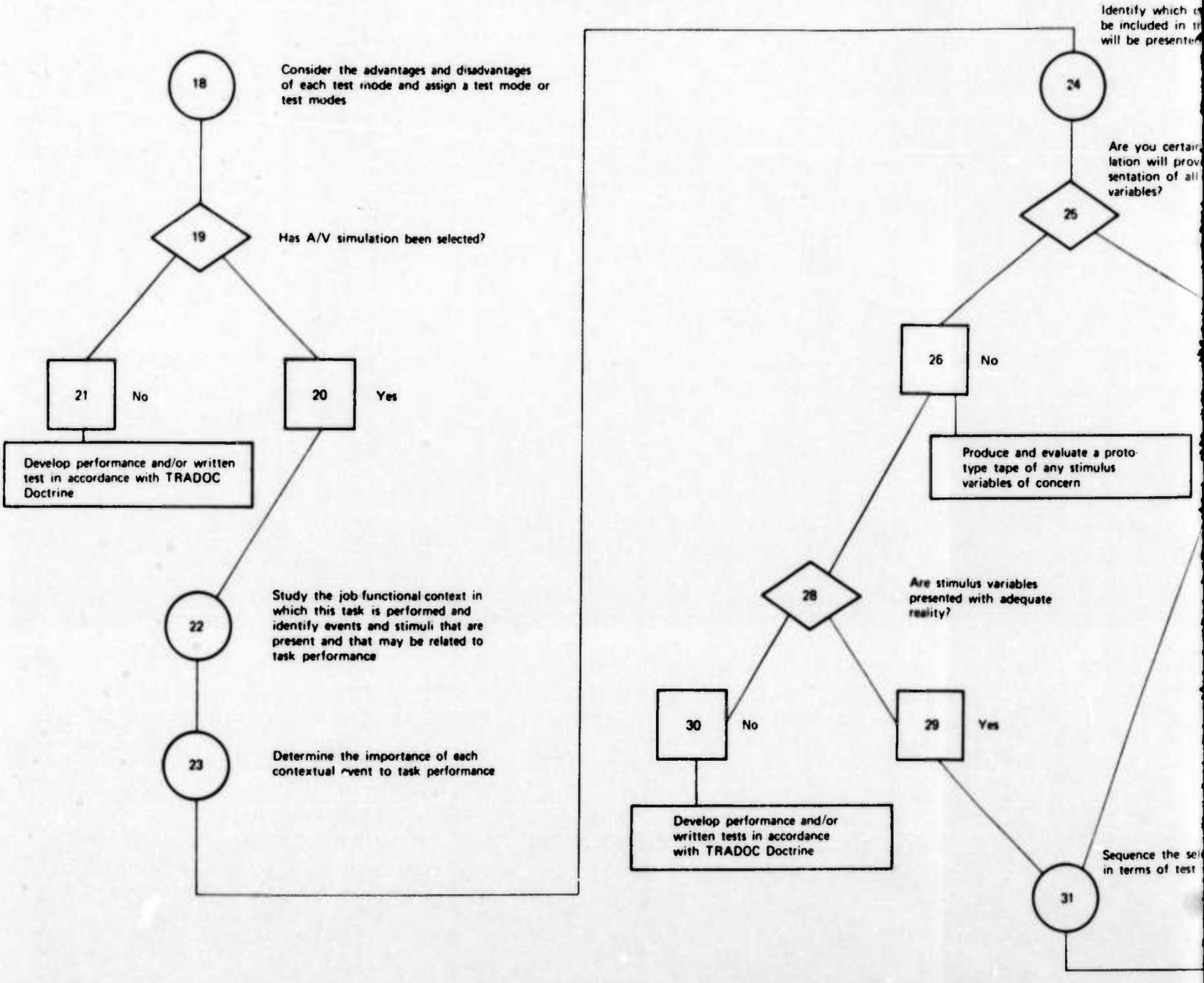


Figure 4. Simulation Algorithm (Sheet 1 of 2)


18
Consider the advantages and disadvantages of each test mode and assign a test mode or test modes
19
Has A/V simulation been selected?
21
No
20
Yes
Develop performance and/or written test in accordance with TRADOC Doctrine
22
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
23
Determine the importance of each contextual event to task performance
24
Identify which
be included in t
will be presente
Are you certai
lation will prov
sentation of all
variables?
25
26
No
Produce and evaluate a proto type tape of any stimulus variables of concern
28
Are stimulus variables presented with adequate reality?
30
No
29
Yes
Develop performance and/or written tests in accordance with TRADOC Doctrine
Sequence the sel
in terms of test
31

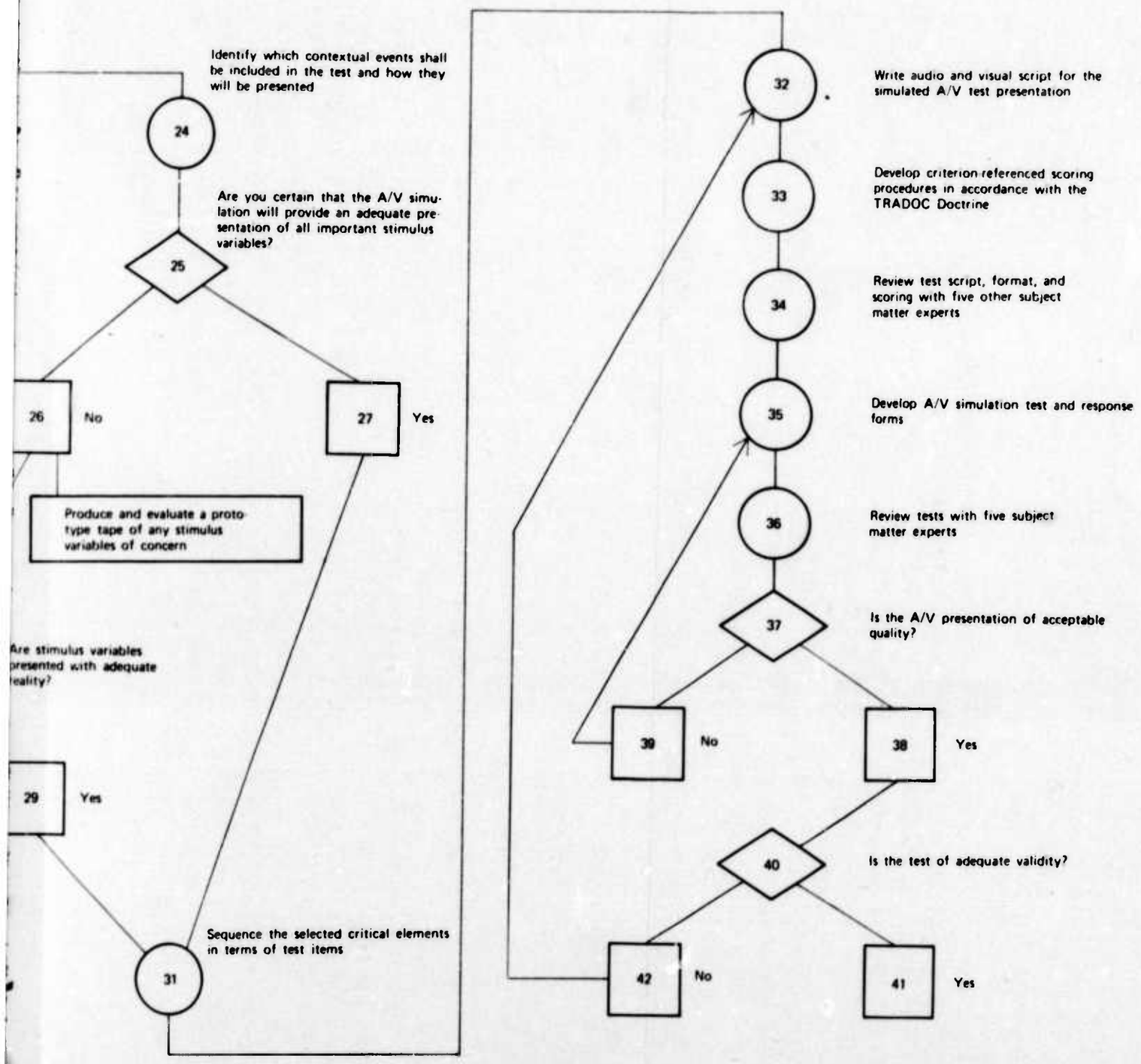


Figure 4. Simulation Algorithm (Sheet 2 of 2)

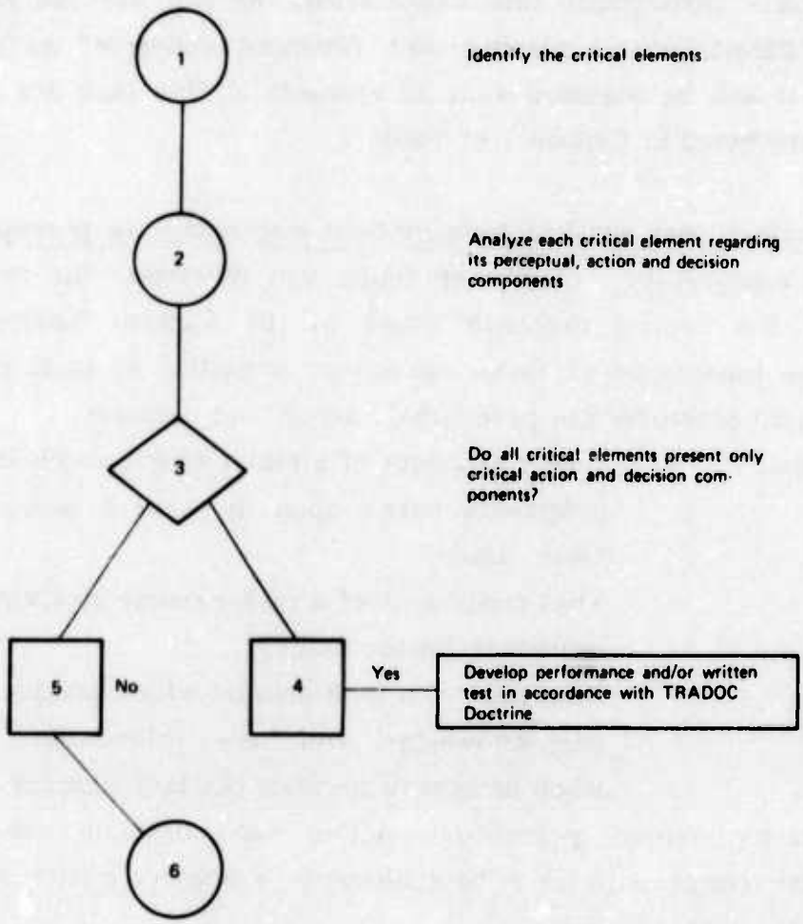


Figure 5. Preliminary Test Mode Selection

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. Critical elements should be identified 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 1.

(2) Analyze each critical task element regarding its perceptual, action, and decision components. This step helps 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)

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

(1) When an error occurs in the performance of this task element, is it often because one fails 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 an error occurs in the performance of this task element, is it often because one fails 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 an error occurs in the performance of this task element is it often because one is misusing knowledge or information?

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

TABLE 1. ELEMENT ANALYSIS TABLE

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

	Critical Element Column 1 (Step 1)	Critical Components Column 2 (Step 2)	Stimulus Variables Column 3 (Step 6)
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	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	

EXAMPLE - The critical components for the elements of the task to "Direct/Control placing and finishing concrete" are listed in Column 2 of Table 1. 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 (i.e., deciding what they should do and observing their performance).

(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.

(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 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.

[4] YES. If the answer is yes a written and/or a performance test should be developed for the task.

[5] NO. If the answer is no, 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 (6.)

EXAMPLE - Column 2 of the Element Analysis Table (Table 1) 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.

DETERMINATION OF TEST REALISM

General

A task/element that has perceptual content is a candidate for A/V testing; but, it is necessary to determine whether the A/V mode can present the task/element in a realistic manner. Figure 6 (Steps 6 through 11) outlines the sequence of operations leading to this determination.

Simulation Algorithm

(6) Identify and list the stimulus variables associated with each critical element. The purpose of this step is to determine specifically 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 responds 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 the Element Analysis Table (Table 1), five stimulus variables have been listed for the sixth critical step - "Direct/Control finishing concrete using a wood float." This example indicates that in performing this step, a person responds to five 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.

(7) Determine the importance of each stimulus variable within each critical element. The purpose of this step is to evaluate the importance of the stimulus variable to the proper performance of the critical element. This is

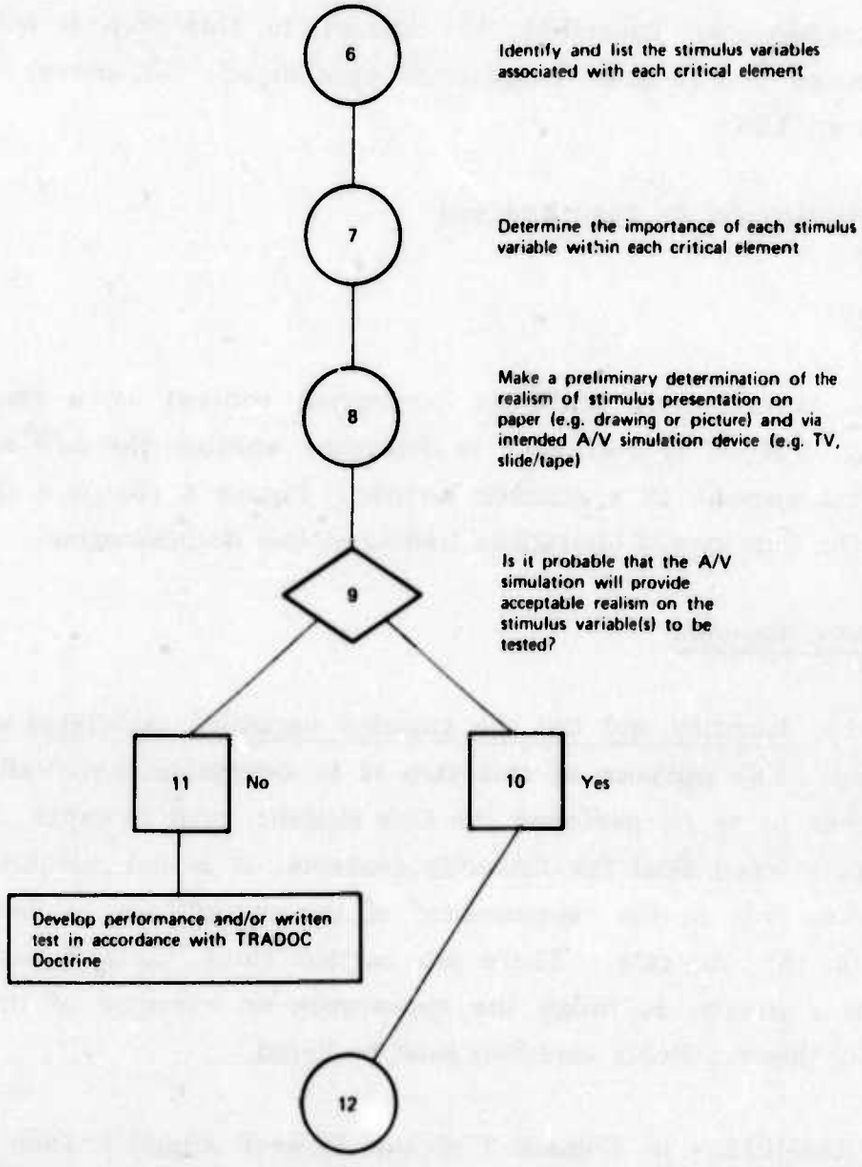


Figure 6. Determination of Test Realism

done because simulation often degrades some aspects of the stimulus and if this step is performed properly, it helps anticipate the effect of any degradation.

In accomplishing this step, the subject matter expert should (1) consider all of the information which is provided by all of the stimulus variables which are typically present when this task element is performed, and (2) assess how often each stimulus variable provides unique information that is essential to proper performance of the task element.

TABLE 2. DETERMINING THE IMPORTANCE OF EACH STIMULUS VARIABLE ASSOCIATED WITH A CRITICAL ELEMENT

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

Stimulus Variables (Step ⑥)	Unique * (Step ⑦)	Essential * (Step ⑦)	Importance * (Step ⑦)
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

A stimulus variable is obviously 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 - Column 3 of the Element Analysis Table (Table 1) lists five stimulus variables associated with the sixth critical element. A judgment is made of the importance of each stimulus variable. 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. 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. Table 2 provides an example of the rating of five stimulus variables.

(8) 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 estimate whether a critical element is best tested via an A/V device or a paper-and-pencil format. A subject matter expert and a training media expert must work together. They will look at each stimulus variable identified in Step (6) and determine the more realistic mode of presentation. Often 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. These are: (1) when the observation of motion or of constantly changing physical characteristics of an environment is important, (2) when representation of three dimensional relationships is important, and (3) when the coordination of sound and visual stimuli is important.

There are also conditions which will favor a drawing or picture. Small color or brightness 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 3. 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 3. 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 (Step 6)	(2) Importance (Step 7)	(3) Comparison of Test Formats Realism (Step 8)	(4) Preliminary Recommended Presentation of Format (Step 9)
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 Column (4) of Table 3. 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 should provide for greater realism in simulating the behavior which occurs as these stimulus variables are observed.

◇9◇ Is it probable that the A/V simulation will provide acceptable realism in presenting very important stimulus variables? 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 provides 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 3. Column 2 contains the judgments made earlier in Step (7) and Column 3 contains the judgments made in Step (8).

[10] YES. If the answer is "YES" an A/V format should be considered. However, further analysis is required.

[11] NO. If the answer is "NO" none of the important stimulus variables can be presented by A/V with acceptable realism, and performance or written testing is necessary.

EXAMPLE - The entries in Columns 2 and 3 of Table 3 indicate that four of the five stimulus variables are 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, or three, a "YES" answer should still be given. In a later step (Step (18)) an A/V test format will be rejected if it is judged to be too limited in scope.

DETERMINATION OF RESPONSE REALISM

Background

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 responses are adequately similar. Figure 7 (Steps 12 through 17) shows the sequence of actions necessary to make this determination.

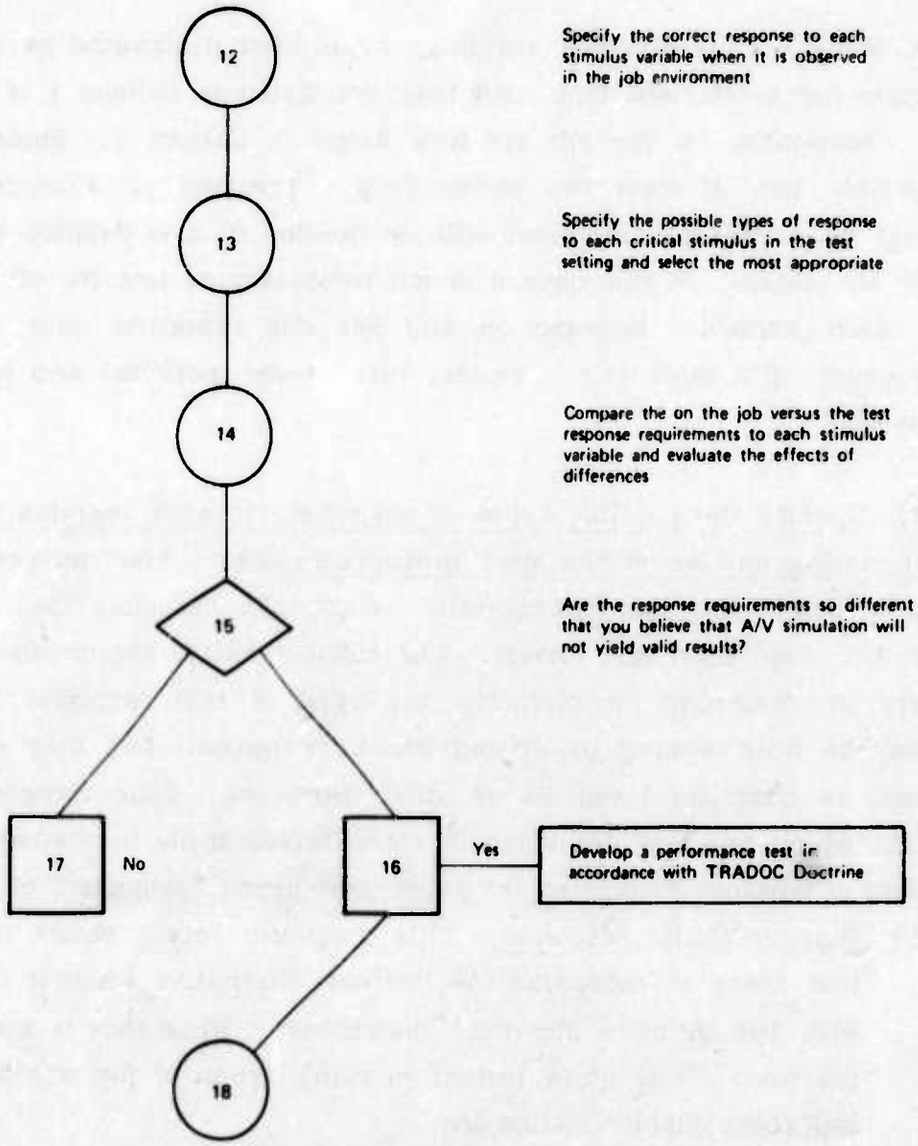


Figure 7. Determination of Response Realism

Simulation Algorithm

(12) Specify the correct response to each stimulus variable when it is observed in the job environment. 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. In doing this, subject matter experts should list each stimulus variable or group of variables that are appropriate for TV simulation and identify the correct job response to that stimulus.

EXAMPLE - Four stimulus variables have been designated as appropriate for a televised test, and they are listed in Column 1 of Table 4. Responses on the job are now 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.

(13) Specify the possible types of responses to each stimulus variable in the test setting and select the most appropriate one. 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. Simulation tests may be built around paper-and-pencil responses, but they may involve responses on computer terminals or other hardware. Four categories of test responses which are now discussed in some detail, apply to response selection regardless of whether responses are paper-and-pencil, computer, etc.

(1) Multiple-choice response - This 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 are:

(a) The test taker must simply recognize the correct response from a limited number of alternatives.

(b) The test taker is 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

TABLE 4. COMPARISON OF RESPONSE REQUIREMENTS TO STIMULUS VARIABLE

(Step 6) Stimulus Variables	(Step 12) Response on the job	(Step 13) Response possibilities on TV Test	(Step 14)	(Step 17) **Recommended Presentation Format
1. Presence or absence of swirls in concrete	a. Direct worker to remove swirl	*a.(1) Report observing a swirl on answer sheet – unalerted identification response	Adequate similarity	
a. If present		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.
** Recommend presentation format for entire critical element – TV

obvious possible actions are all possible at the same point in time. The format is appropriate 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 process, such as when to stop vibrating concrete.

(3) Unalerted identification response - This type of response differs from the multiple-choice and alerted two-alternative response in a very important respect. The test taker must identify an event without being told when it may occur. 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 an 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 4 (Step (13)) 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 4), and the types of test responses which have been discussed in the preceding paragraphs. The recommneded test response to Stimulus Variables 1, 2 and 3 in Table 4 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 nd 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.

(14) 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 largely 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 vs 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 doing the job and then evaluate the 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 dissimilarity is most likely to arise when attempting to test an action component via audio-visual simulation.

(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 of such job skills which only requires a response to one isolated task

may not provide valid results because of the discrepancy between the job environment and the test environment.

(3) Differences in difficulty of job and task responses - This situation is similar to the preceding. 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 often occurs when 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 (13) 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.

EXAMPLE - The judgments called for in Step (14) are listed in Column 4 of Table 4. 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 of a supervisor is a "critic" response as is the test response.

◇15◇ Are the response requirements so different that you believe that A/V simulation will not yield valid test results? This question is asked because you now have sufficient information to determine whether to develop an A/V simulation test.

[16] 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.

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

EXAMPLE - Column 4 of Table 4 indicates that the test responses to all four stimulus variables are adequately similar to provide valid test results. The answer to Step ◇15◇ is therefore "NO."

FINAL TEST MODE SELECTION

Background

Up to this point it has been determined that stimulus and response variables allow for an A/V simulation test. At this time it is necessary to decide whether or not to use the A/V test mode. In making this decision one should consider factors such as cost, the feasibility of testing selected stimulus variables and not testing others, or of testing different stimulus variables by different test modes. Figure 8 (Steps 18 through 21) provides the sequence of actions followed to select the proper test mode.

Simulation Algorithm

(18) 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.

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:

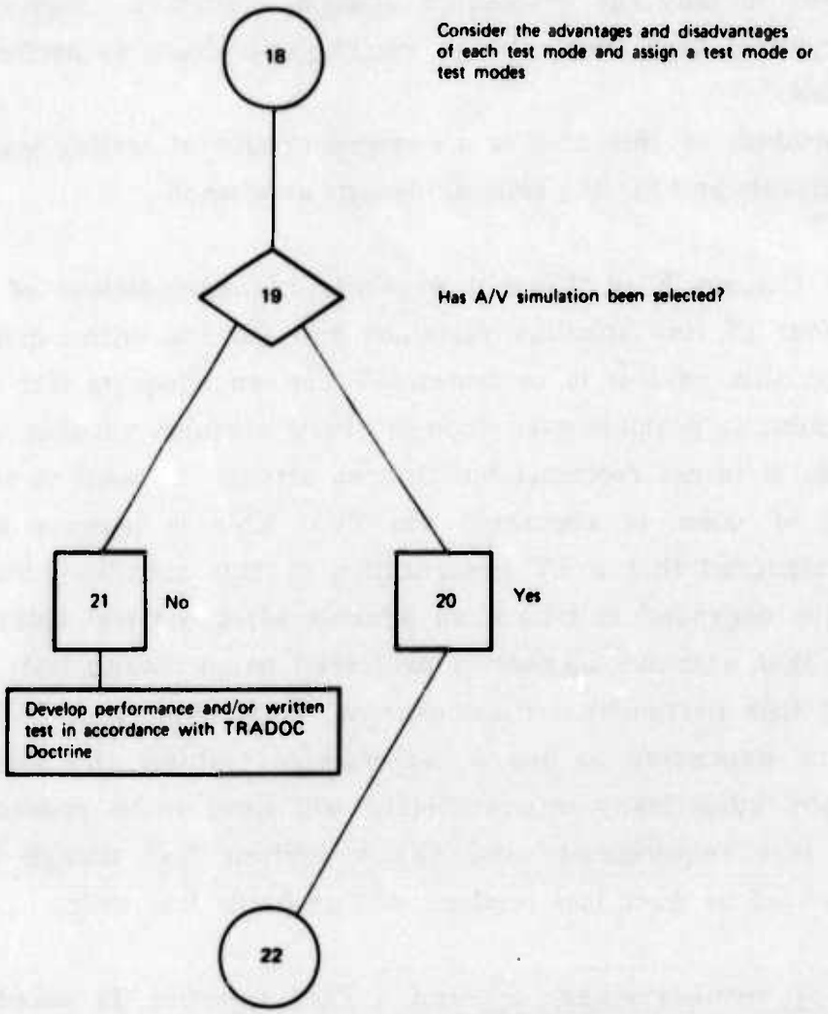


Figure 8. Final Test Mode Selection

(1) Select the single test mode that is appropriate to the most highly critical elements and stimulus variables and test as many critical elements as possible via that mode. The mode which is used provides a valid test for all components which are included.

(2) Test critical elements or stimulus variables via different testing modes. For example, if performance tests are used, neither written nor A/V simulation are typically 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 as easily 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 4 presents recommendations of TV tests for four of five stimulus variables and for the entire critical element. In this case it is recommended that an adequate test of a critical element is possible even though every stimulus variable isn't tested; also, 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 validity. No test of that stimulus element is preferred to an invalid 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.

◇19◇ 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.

[20] YES. If the answer is "YES", the decision is to develop an A/V simulation test and continue with test development.

[21] NO. If this answer is "NO", an A/V test is determined inappropriate.

REPRESENTATION OF JOB CONTEXTUAL STIMULI

Background

Earlier steps have identified stimulus variables which are part of the critical tasks and specific responses. A/V simulation enables us to enhance validity by presenting some of the job environment in the simulated test.

The following steps are presented to determine how or if job related stimuli and events can be adequately presented.

Figure 9 (Steps 22, 23 and 24) depicts the sequence of actions necessary to determine this representation.

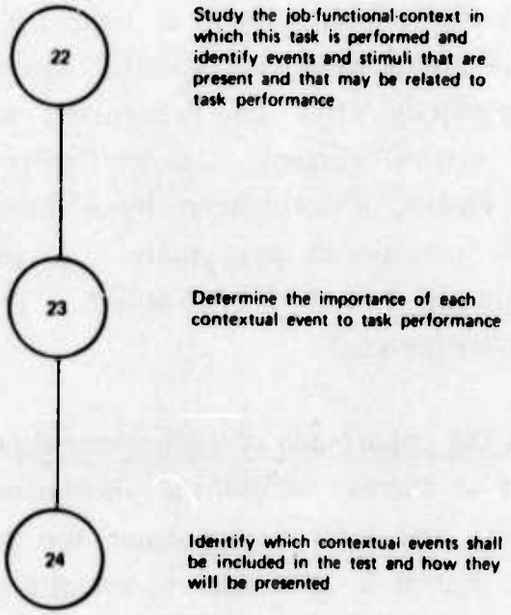


Figure 9. Representation of Job Contextual Stimuli

Simulation Algorithm

(22) 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. In performing a task, one is influenced by factors which are not an integral part of the task itself. Examples are factors such as noise, quality of lighting, disturbances, temperature.

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 - 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. The critical element, "Direct/Control Finishing Concrete Using a Wood Float", is influenced by a number of variables that have not been considered previously. A partial list has been entered in Columns 1 and 2 of Table 5. This list provides the basis for the following step.

(23) Determine the importance of each contextual event to task performance. This step is to further define the simulation requirements. It is not necessary to simulate the total environment for valid use of simulation in training or testing, but it is beneficial to include stimuli that heavily influence one's performance. The same basic question presented earlier, in Step (7) is repeated in this step. In this step it is necessary for a subject matter expert to review the list of contextual events which were developed in Step (22) and: (1) consider all of the information which is provided by all of the contextual events which are typically present when this critical element is performed and (2) assess how often does each 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 it's 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 5. EVALUATION OF THE JOB-FUNCTIONAL-CONTEXT

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

(1) Contextual Event (Step 22)	(2) Contextual Stimuli (Step 22)	(3) Criticality of Contextual Event (Step 23)	(4) Inclusion of Event In Test (Step 24)	(5) Stimuli Selected for Presentation (Step 24)
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 5 are reviewed. A judgment is made of the importance of each event in performing the critical element. In Column 3 of Table 5 this judgment is recorded. All events were judged to be very important because it was assumed that each provides unique information that is essential to proper performance.

(24) 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.

EXAMPLE - In this example all of the contextual events which are identified in Step (22) are judged very important in Step (23) . 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 5 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

Background

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 are shown in Figure 10.

Simulation Algorithm

<25> 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 simulation 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.

[26] NO. If the answer is "NO", a prototype tape that will enable a judgment of the adequacy of simulation of specific stimulus variables should be developed.

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.

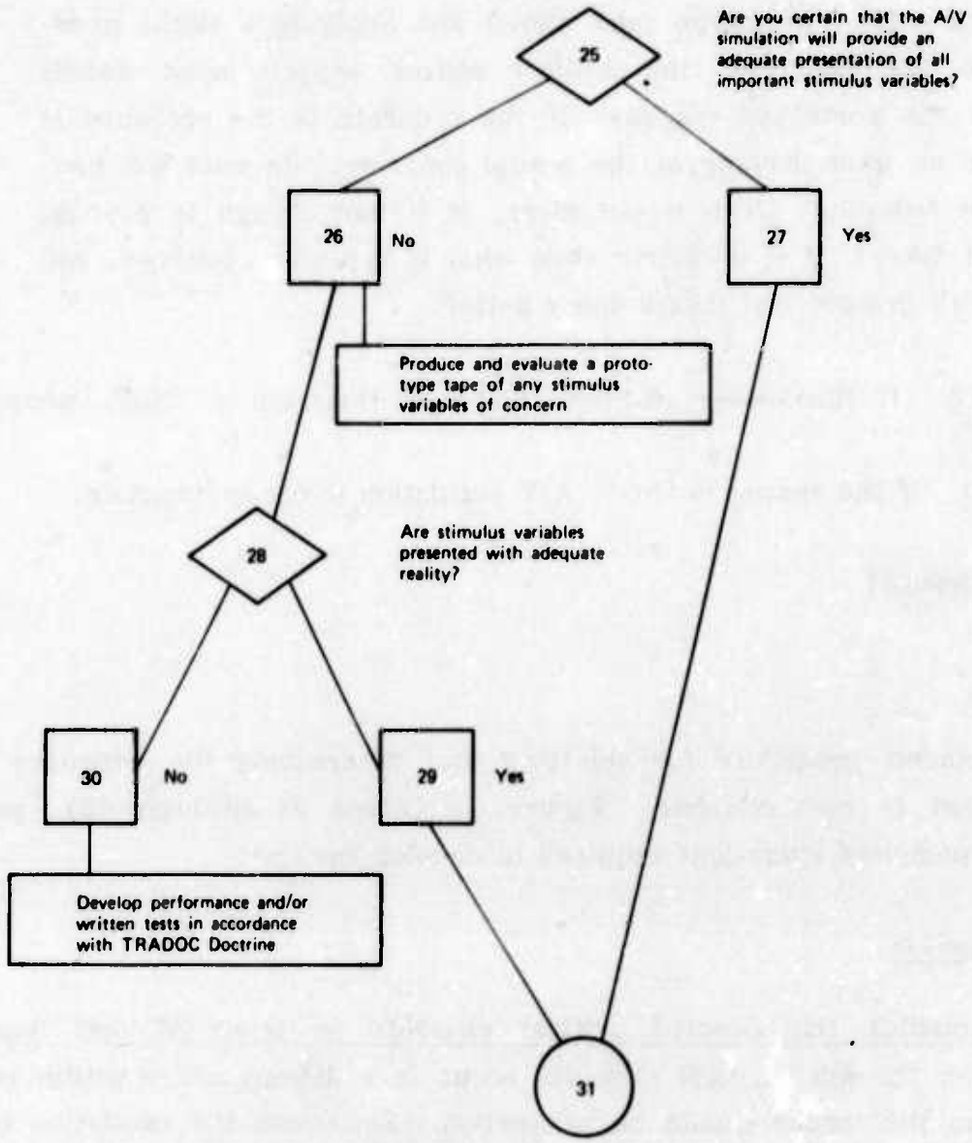


Figure 10. Final Assessment of Presentation Realism

27 YES. If the answer is "YES" proceed to Step (31).

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

[29] YES. If the answer in Step [27] or this step is "YES", proceed to Step (31).

[30] NO. If the answer is "NO", A/V simulation is not appropriate.

TEST DEVELOPMENT

Background

The detailed procedure for selecting and determining the adequacy of A/V simulation is now complete. Figure 11 (Steps 31 through 42), presents the sequence of operations required to develop the test.

Simulation Algorithm

(31) Sequence the selected critical elements in terms of test items. 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.

(32) Write audio-visual script for the simulated A/V test presentation. In Step (31) the critical elements to be tested were sequenced; now it is necessary to specify the exact audio and visual content of the test. 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:

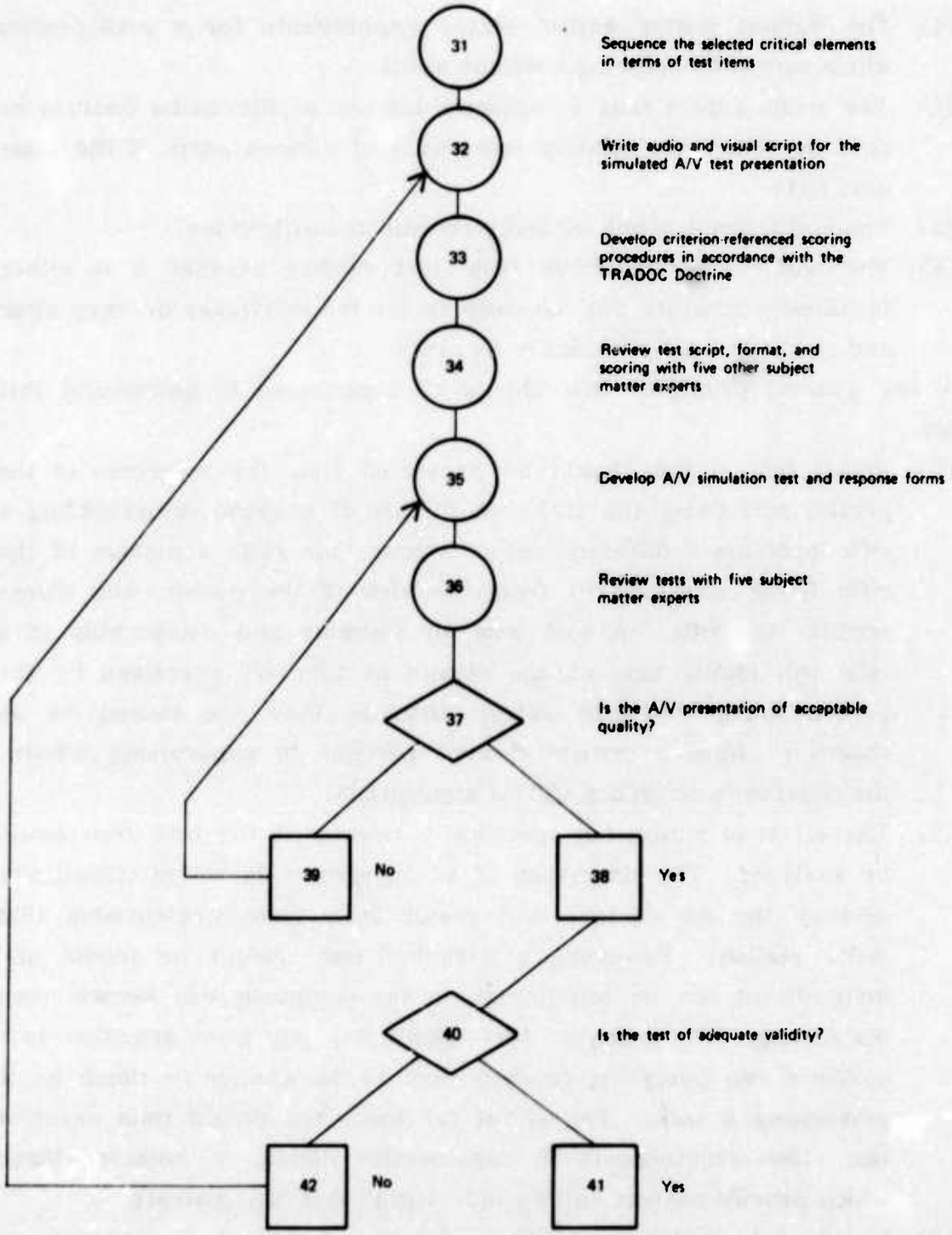


Figure 11. Test Development

(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) Visual test stimuli should be presented 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 should be analyzed. The elimination of all apparently unrelated stimuli may destroy the job context and result in a video presentation that lacks realism. However, 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. The appended "Television Production Considerations," should aid you in performance of this step.

(33) Develop criterion-referenced scoring procedures in accordance with the Manual for Developing SQTs. 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.

(34) Review test script, format and scoring with five other subject matter experts. Before developing the test, 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 element. 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) 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 involve continuing review and test revision until all subject matter experts agree the test should provide a technically accurate and valid measure of the critical elements.

(5) Refer to the Appendix for more detailed guidance.

(35) Develop A/V simulation test and response forms. Audio and video recording and editing must be accomplished by personnel who are skilled in production techniques and procedures. A subject matter expert must 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.

(36) Review tests with five subject matter experts. The subject matter expert should be given the test, just as it is to be given to other soldiers. Extra instructions or background must 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. All errors on the test made by subject matter experts should be analyzed. 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. All subjective comments should be recorded and studied. Favorable comments are nice to hear but pay more attention to the unfavorable. Remember, the ultimate goal is a well developed test.

<37> Is the A/V presentation of acceptable quality? This step occurs simultaneously with the conclusion of Step (36) . Consider this question as you analyze the test results and comments obtained from subject matter experts.

[38] 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.

[39] 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 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 (35) .

<40> Is the test of adequate validity? Earlier steps established the content validity of the A/V test based upon the judgments of subject matter experts who possess the knowledge and skill of the MOS. 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.

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

[42] NO. This answer will be selected if the test is not determined valid and the test developer must go back to Step (32) .

Appendix

TELEVISION PRODUCTION CONSIDERATIONS

APPENDIX
TELEVISION PRODUCTION CONSIDERATIONS

INTRODUCTION

If the test developer is considering the use of an audio-visual simulation test, it will be extremely helpful to involve a media specialist in the test development process at the earliest opportunity. To be most effective, the media specialist must become very familiar with the subject matter to be tested. The media specialist should become involved prior to the script-writing stage. It will be his job to suggest ways in which the test items can be most effectively simulated. Revisions to videotaped productions can be time consuming, costly, and in some cases, impossible. In many cases, videotape production units within the Army run on very tight schedules and can require as much as three months lead time on a production.

The most compelling reason for early, close coordination between the test developer and the media specialist is that each will fill the gaps in the other's knowledge. Very often the test developer will over-estimate the capabilities of the videotape production unit. Most people are acquainted with videotape only through what they see on their home sets. The attitude seems to be, "I saw this done on Monday Night Football," or "on such and such a commercial," and therefore feel the effects can be easily duplicated for their test. What the developer must realize is that network television represents a massive investment in time, talent and equipment. Although the talent may be arguable, the equipment used in American network television productions is th newest, most advanced and most expensive in the world. What looks easy on the TV set in your living room, may simply be impossible for an industrial or Army production unit. A qualified media specialist will be able to point this out early, if he's made aware of what's happening. In the event that this early coordination is impossible, this appendix briefly discusses some do's and don'ts of television production.

AUDIO-VISUAL PERFORMANCE SPECIFICATIONS

PURPOSE

This performance specification provides minimal criteria for the objective review of audio-visual programs developed for testing purposes, and constitutes the basic quality assurance standard against which new and revised audio-visual materials should be evaluated for acceptance.

SCOPE

This performance specification is designed specifically for application to videotape material, though the same elements may be relevant to the evaluation of material presented in a film format. These criteria should not be applied to materials presented solely in an audio format or solely in a video format. There is an interaction of audio and video that is required for television and film, and these are defined with that requirement in mind.

DESCRIPTION

The Performance Specification Checklist and abbreviated Definition of Terms were constructed for use by the test developer who is not necessarily schooled in videotape production techniques and terminology. As such, this is not meant to be a comprehensive list of every element involved in the production of an audio-visual presentation. Rather it is a list of crucial elements which should be taken into consideration in evaluating a program.

Before reviewing any simulation testing, the test developer should familiarize himself thoroughly with the checklist and definition of terms. The program should be viewed a minimum of two times to allow the reviewer to respond on both a general level and a specific level. While viewing the program, he should record noteworthy details in the "Comments" column. Then, after viewing the program completely, he should assign numerical ratings to each of the items listed.

PRODUCTION ELEMENTS

An audio-visual program can be divided into a number of interrelated production elements, each of which makes a contribution to the whole. The Performance Specification Checklist identifies those elements in an effort to provide a standard approach and format for evaluating training materials. The following is a detailed explanation of these individual elements.

Audio

Audio refers to the sound accompanying the "video" or visual portion of a program. Audio includes the selection of music, use of ambient sounds and/or sound effects and recording of voices involved.

Ambient sounds are natural noises recorded in the environment; e.g., the whirling sound of a machine, the blast of a siren. Sound effects are any sound, other than music or speech, artificially reproduced to create an effect; e.g., the sound of footsteps, the screech of car brakes. Sound effects can add a dramatic touch to a program, but they should be used only if there is a definite need for them. For example, if the scene shows a key punch operator at work and there is no narration during that segment, then the sound of the key punch could be heard.

The voice of the narrator should fit the subject matter. Whether male or female, the narrator should have a voice with pleasant tonal and rhythmic qualities. Pitch and inflection should underscore the meaning of the script. The pace of the narration should be steady and easy to follow. In most cases, the narrator plays a large part in how the production ties together.

Camera

Everything displayed on the video screen is seen through the eye of a camera. How an individual, equipment, or scene comes across is largely a function of how the camera is manipulated; so camera movement, focus, and picture composition are of the utmost importance.

Camera movement includes the use of a zoom lens or the actual physical movement of the camera to change the visual perspective. The transition from one position to another should be smooth and even. Changing camera

angles can give a production variety. But if changes are made too frequently and without a reason, the program will be disjointed and confusing.

A picture is in focus when the image appears sharp and clear on the screen. The main subject should be visually well-defined. Occasionally, a fuzzy image is used for effect. If it is meaningful, it is permissible. However, this is rarely the case.

The composition of a shot takes many factors into consideration. The subject (props or people) should be arranged to complement the background and setting. The elements in the picture should be well balanced within the frame of the shot to create a pleasing visual image. The best camera angle is the one that shows the subject most clearly.

Special Effects

Special effects are the miracles of the electronic medium. This category includes electronic effects, optical effects and mechanical effects ranging from superimpositions to split screens and visual insertions.

The most important consideration is whether or not the special effect is necessary to the overall meaning of the program. It should serve a specific purpose whenever it is used. Special effects should be used sparingly. If they are over-used, they lose their uniqueness and effectiveness. In some cases, special effects can add to the production time and increase the difficulty of editing. Special effects used when not absolutely necessary will likely distract the viewer.

Lighting

Under the best conditions, lighting can help to create the illusion of time, locale, mood and style of production. The subject should be lit sufficiently with a good distribution of light and shadow to show up all details clearly. Colors should reproduce well.

Cameras need a specified amount of light to produce an acceptable electronic picture, and color cameras are particularly sensitive to lighting conditions.

Shooting on location often introduces lighting problems which cannot be controlled. For instance, the practical exercise buildings at Fort Leonard Wood, MO, have extremely high ceilings, assorted colors of fluorescent lights which interact with the camera and limitations on the kinds of additional lighting which may be used to bring the light levels up to the necessary intensity. These conditions are far from ideal for accurate reproduction of a scene, and this should be taken into consideration when evaluations are made.

Scenery and Props

Scenery and properties (i.e., props) are the physical setting and objects which help the viewer identify the locale for a program. Typically, audio-visual material prepared for SQTs would be shot on location, eliminating the requirement for fabricating the set. Individual pieces of hardware become props. The major criterion for selecting and evaluating scenery and props is whether they aid in interpreting the production, rather than distract from the main focal point.

Graphics

Graphics include two-dimensional visuals such as title cards, illustrations, maps, charts, slides, and photographs; i.e., any art work used in a program.

The type of graphics to be used is dictated by the subject matter. If the subject is a complex one, it may be most appropriate to present it with a series of simple illustrations that can characterize the interaction of the major points. If mood or image is to be conveyed, a caricature may be most suitable. If accuracy is the primary concern, a photograph or slide may be a good technique. The choice is determined by the visual aid which will convey the message best.

The relatively small, proportionally fixed screen size of the television set demands maximum clarity for all graphics. Art work should be precise and uncluttered. The proportions of the viewing screens are three-by-four, that is, the picture is always three units high and four units wide. All picture information prepared for video must be contained within the three-by-four aspect ratio.

The preparation of graphics also requires attention to color aesthetics and to some technical aspects of the color camera. The color combination should be appropriate for the intended communications. Sharp brightness contrast is essential, since some sort of written information is almost always used. Colors and designs that detract from the purpose of the art work should be avoided.

Performing and Acting

In most cases, the people performing in videotaped simulation performance tests have had little if any experience before a camera. They should not be expected to give performances that meet professional standards, but they should be aware of the camera and they should relate to it. It is probably unwise to use one performer throughout the production. If the test requires revision, which is most likely, there is a chance that the performer may not be available to you again. This could mean reshooting the entire production. If they are demonstrating something, they should hold the object or perform the procedure in such a way that the camera can follow along. One of the major problems associated with demonstrations is the tendency of the talent to block the camera's view; and if the camera cannot see the demonstration, neither can the viewer.

Production/Direction

Production and direction involve the coordination of all production elements blended to create the final product. The important aspects of this category include pace, and editing of audio and video.

Pacing is the overall rate at which the material is presented. The subject matter, the degree of difficulty, and the knowledge levels of the target audience determine the speed at which the information is covered. Audio editing and video editing refer to the act of tying these segments together at the prescribed pace. The visual segments should be arranged in a logical order with a minimum of electronic disturbances. The sound should be well synchronized with the video to complete the message.

Script

The script should be organized in a logical manner with good attention to details of major importance. A script should be evaluated in terms of clarity and flow. The language should be simple and well defined; the images carefully selected for maximum effect. Audio and video should blend well to convey the total meaning. It is not always necessary to show everything that is said or say everything that is shown. Audio and video not only reinforce one another through repetition; they can expand on the message by highlighting different aspects of the subject.

The flow of the script is determined by the pace and continuity of shots. The transition from one concept to the next should move smoothly and naturally. The viewer should be able to follow the program as it proceeds to a logical conclusion.

Interest Factor

A program must capture the attention of the viewer before it can make a valid simulation test. The material should be presented in an interesting manner to involve the audience.

General Evaluation

The general evaluation category represents the culmination of responses to all the other production elements. This is not simply a numerical average of the other ratings, for a program could score low on a number of items and overall still be very successful in achieving its purpose. Rather, this is a subjective reaction to the presentation; i.e., a personal analysis of the value of the production.

SUMMARY

The Performance Specification Checklist (Figure 12) and accompanying Definitions of Terms is one approach to the difficult task of reviewing and evaluating audio-visual materials prepared for testing.

As outlined, numerous production elements intermix to create the final product. Each of these elements is important in and of itself; consequently, it should be taken into account when judgment is made. However, the degree of importance of each element will vary from production to production. For example, one program may have only opening and closing graphics, while another relies heavily on graphics to convey ideas in the text of the program. In the first case, graphics would play a much smaller part in the overall effectiveness of the program than they would in the second case.

It is the responsibility of the test developer to apply variable weighting to these elements as they pertain to individual programs. In this way, the performance specification can and will be a flexible set of guidelines for reviewing audio-visual materials developed for testing.

PRODUCTION TITLE:			RATING low					high
PRODUCTION ELEMENTS		COMMENTS	1	2	3	4	5	6
A. AUDIO								
B. CAMERA	Camera Movement							
	Focus							
	Composition of Shots							
C. SPECIAL EFFECTS								
D. LIGHTING								
E. SCENERY & PROPS								
F. GRAPHICS	Selection of Art Work							
	Clarity							
	Size							
	Color							
G. PERFORMING & ACTING								
H. PRODUCTION/DIRECTION								
I. SCRIPT	Clarity							
	Flow							
J. INTEREST FACTOR								
K. GENERAL EVALUATION								

Figure 12. Performance Specification Checklist

DEFINITION OF TERMS

AUDIO

Selection of sound effects, ambient noise and voice. Example: Is the ambient noise appropriate for the subject?

CAMERA

Camera Movement - the transition from one shot to another. Example: Did the camera zoom in with one flowing motion or was the movement jerky?

Focus - the clarity of each shot. Are images fuzzy or well-defined?

Composition - the way the subject matter in the shot is arranged. Are elements in the picture well balanced within the shot? Is the subject matter shot from the best angle?

SPECIAL EFFECTS

The electronic, mechanical or optical "tricks" used to highlight aspects of the program. Do they serve a purpose or are they merely gimmicks?

LIGHTING

The distribution of light and shade in a shot. Was lighting sufficient to show up all details clearly? Did the light levels remain relatively constant throughout the production?

SCENERY AND PROPS

Were background and props realistic and pertinent? Did they aid in interpreting the production?

GRAPHICS

Selection of Art Work - is the type of graphic used most appropriate to the program? Example: Was a caricature used when a photograph would be more meaningful?

Clarity - precise, uncluttered art work.

Size - use of the proper 3 x 4 aspect ratio required for television.

PERFORMING AND ACTING

Relationship to the Camera - how the talent reacts to the camera. Is the talent at ease on camera?

PRODUCTION/DIRECTION

Pace - the overall tempo. Is the rate at which the material is presented suitable for the subject matter?

Video Editing - electronically arranging visual segments.

Audio Editing - synchronization of sound tracks with video presentation.

SCRIPT

Clarity - the language and organization of the script. Is the language easily understood? Do the details fall in a logical sequence?

Flow - the pace and continuity of shots. Does the program move smoothly? Is the response time adequate?

INTEREST FACTOR

The degree to which the program involves the audience. Is the program interesting to watch or is it dull and boring?

GENERAL EVALUATION

Over-all response to the program.