AD NUMBER

AD011744

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; 31 JAN 1953. Other requests shall be referred to Office of the Adjutant General, Personnel Research Branch, Washington, DC 20310.

AUTHORITY

ARI ltr dtd 13 Nov 1979

PRR REPORT NO. 991
THE DEVELOPMENT OF A SOUND MOTION PICTURE PROFICIENCY TEST

By
C. R. Carpenter,
Principal Technician
Instructional Film Research Program
The Pennsylvania State College
State College, Pa.

A CONTRACT RESEARCH REPORT TO

DEPARTMENT OF THE ARMY

THE ADJUTANT GENERAL'S OFFICE

PERSONNEL RESEARCH AND PROCEDURES DIVISION

PERSONNEL RESEARCH BRANCH

PRB REPORT NO. 991

THE DEVELOPMENT OF A SOUND MOTION PICTURE PROFICIENCY TEST

by

C. R. Carpenter, Director
L. P. Greenhill, Program Coordinator
W. F. Hittinger, Research Assistant
E. P. McCoy, Research Assistant
C. J. McIntyre, Research Associate
J. A. Murnin, Research Assistant
R. W. Watkins, Research Assistant

Instructional Film Research Program
The Pennsylvania State College
State College, Pennsylvania

FINAL REPORT 31 January 1953

DEPARTMENT OF THE ARMY
Project No. 29549100 Subtask 79
RESEARCH CONTRACT NO. DA-49-083 OSA-393
PERSONNEL RESEARCH SECTION PR 4979

PRB Reports are technical reports. They are intended primarily for research agencies in the Armed Forces as a means of guiding further research in the area of human resources. As research findings accumulate and suggest official action, recommendations are made separately to appropriate military agencies. Information of more general interest is presented in the Foreword to this report.

FOREWORD

When it is appropriate to describe a particular work assignment with words, when words can be used to create the job situation, then effective paper-and-pencil tests of job proficiency can be developed. However, there are many jobs requiring manual skills, the handling of equipment, the repairing of machines -- the man is required to do something with something, not merely to talk about it. Likewise, the examiner may be less interested in the man's ability to tell how to repair a given piece of equipment than in knowing whether he is effectively trained to actually do the job. Paper-and-pencil tests are adequate to satisfy the former interest, but the question has been raised of whether they are adequate for real, live performance.

This report describes research to explore the potentialities of sound-motion pictures as a method of testing proficiency in jobs requiring complex muscular skills and adjustments. Can a sound-motion picture test be constructed for measuring job performance that is a more realistic, more reliable, and more accurate predictor of success than a paper-and-pencil test?

A sound-motion picture test was constructed for measuring the proficiency of track vehicle repairmen. The job for which men are beginning their training requires them to repair tanks and related equipment. The film test emphasized the sights and sounds to which a repairman must attend in order to find the causes of the trouble -- clues a sound-motion picture can present. The film test was found to be highly reliable and to yield much the same results as the paper-and-pencil test already in use for this training course. However, in comparison with instructors' ratings of the proficiency of the trainees, the film test was not much better than the current paper-and-pencil test.

There is still the possibility that a sound-motion picture may be a better testing method for some purposes, but in this research the method was not proven to be superior to methods already available.

ACKNOWLEDGMENTS

The Instructional Film Research Program of The Pennsylvania State College would like to acknowledge the help it received from a number of sources in constructing and proving the Motion Picture Proficiency Test.

Many valuable suggestions were made from time to time by members of The Personnel Research Branch, Office of The Adjutant General. Much advice and day to day material help was given by numerous people at The Ordnance School. In particular we should like to acknowledge the help of Major Gordon Clyde and Mr. H. E. King of the Automotive Branch.

The following members of the faculty at The Pennsylvania State College participated in the research: Dr. Philip Ash, Dr. Nathan Jaspen, Dr. Bruce Moore, and Dr. Kinsley Smith. A portion of the report was reviewed by Dr. Albert K. Kurtz; Dr. Harold E. Nelson narrated the film commentary. The College Motion Picture and Recording Studio assisted in the production of the film test, and special mention should be made of the photographic work of Mr. D. P. Duvall.

TABLE OF CONTENTS

I. THE PROBLEM

Sound motion pictures are used extensively for training. Their characteristics and potentialities for realistically representing problem situations, however, appear to offer possibilities for developing and applying the medium as an instrument for testing and evaluating the results of training.

The need to develop more discriminating tests of achievement and aptitude during World War II led to the establishment of the Psychological Test Film Unit by the U. S. Army Air Force in 1943. This unit, under the command of then Lt. Colonel James J. Gibson, appears to have been the first systematically to explore the potentiality of motion pictures as a testing instrument (4). Beyond that, relatively limited use has been made of the medium for testing and evaluational purposes. Among the functions listed by Gibson as being particularly amenable to motion picture testing are the following:

1. Discrimination of visual motion and locomotion
2. Perception of space and distance
3. Maintaining orientation during locomotion
4. Ability to learn a procedure
5. Ability to react to a changing situation
6. Ability to perform under emotional stress

In training men for military occupational specialties, it is necessary to be able to measure accurately knowledge the trainees have acquired. In the past, two main approaches have been used in gauging this knowledge: (1) verbal tests of information (oral or paper-and-pencil) and (2) work-sample or performance tests. The first suffers from the fact that it is unable to measure physical manipulatory skills and abilities. The latter is time-consuming and expensive, since an instructor can test only one trainee at a time, and materials used may be expensive and non-expendable. Such testing also typically entails an undesirable degree of subjectivity.

The purpose of this research therefore was to determine the feasibility of employing sound motion pictures as a measure of proficiency testing in lieu of performance tests. It was further hypothesized that to the degree that the sound motion picture can be used to represent performance situations more fully, realistically and concretely than can be done with paper-pencil tests, there would be a corresponding advantage in terms of validity favoring motion picture tests over paper-pencil tests. It was also hoped to demonstrate that sound motion picture tests could be developed for a range of tasks, especially of the perceptual-motor type, which might have equal or higher validity and reliability than simulated performance tests. In this comparison, moreover, the film tests might have advantages of being more practical since relatively large groups of subjects could be tested simultaneously, the amount of supervision (test administration) could be reduced, and the testing situations could be better standardized. Finally it was felt that film tests might be made authentic both in photography and sound, eliminating or reducing many of the non-relevant factors and interference involved in field testing.

II. PROCEDURES

A. Selection of an Area of Content for Demonstrating Film Tests

Representatives of the Personnel Research Section of the Adjutant General's Office working with representatives of the Instructional Film Research Program of The Pennsylvania State College formulated the following general criteria which would be used in selecting an area or "task" for which a film test would be developed in order to demonstrate the feasibility of this medium:

1. The area should be of the performance type and involve predominantly overt behavior using equipment or tools which could be photographed, thus making possible the use of the potentialities of sound motion pictures for recording complex situations and action.

2. The task should include possibilities for producing authentic sound items, hence sounds should be crucial cues to some of the performances.

3. The task or area selected should have already prepared testing instruments, methods of grading and, if possible, performance tests with which to compare the sound motion picture tests when produced.

4. The area should have potential validation criteria with respect to the subjects tested.

5. The area selected should be one for which a sound motion picture test would be a useful and practical testing instrument once this had been developed and on which continued research would be possible.

6. The area should be of enough importance to military operations to justify the project to military authorities.

The Personnel Research Section representatives recommended The Ordnance School at Aberdeen Proving Ground as the best possibility of a base of exploration. Three courses of instruction were accordingly considered: the Machinist Course, the Small Arms Repair Course, and the Track Vehicle Repairman Course.

The Track Vehicle Repairman Course was selected because it was understood that:

1. Much overt behavior and gross behavior as well as fine motor skills were involved.

2. Sound cues were crucial in some of the tank engine repair work.

3. Paper-pencil tests, performance ratings and performance problems for solution were administered as standard operating procedures in connection with the course of instruction.

4. Validity criteria as scores on work sample performance tests already existed or would be available.

5. A sound motion picture test would be useful when developed and made available.

6. The Track Vehicle Repairman Course is of critical importance to Army operations, and this subject matter area would justify, it was believed, the relatively extensive sound motion picture test project.

B. Construction of Multiple-Choice Film Test Problems

The motion picture proficiency test was developed over a period of about nine months, from December 1951 through August 1952, in accordance with criteria contained in Appendix L2.

1. Training in Course Content

In December, 1951, basic preparation for the construction of motion picture proficiency test problems was begun by a team of six Instructional Film Research Program personnel.

One member of the team was designated as "continuity man." He enrolled in the Track Vehicle Repairman Course at The Ordnance School, Aberdeen Proving Ground. His function was to take the complete 13-week course, familiarize himself with the course content, the trainees' problems, the temper and characteristics of the trainee population, and to provide for the orientation and integration of the remaining five team members, each of whom took a one to two week segment of the course. Each of these five men became thoroughly familiar with the material in the particular course segment which he had covered, and was responsible for constructing test problems in that area.

During this period of training, investigation, and analysis, the members of the Instructional Film Research Program team took the standard course of instruction, classroom and practicum, read the pertinent manuals, held discussions with instructors and students, studied instructors' lecture notebooks and the Lesson Plans for the course. As the team worked through this phase of the research, they undertook to draft test problems along the way.

2. Test Problem Construction

The intensive period of problem construction, however, began immediately following the investigation and training phase. During a three week period approximately 430 tentative test problems were drafted.

3. First Panel Review

After the 430 problems were drafted, a panel reviewed each item by reading it aloud and discussing it. The panel included the problem-writing team and additional faculty members of The Pennsylvania State College who were skilled in test construction and personnel measurement and experienced in sound motion picture research.

4. Check for Technical Accuracy

The next step was to take the problem-scripts to The Ordnance School and have them reviewed by experts for technical accuracy. This included making certain that there was one and only one right answer, that the alternatives were reasonable and, apart from their not being correct answers, were "attractive."

5. Technical Panel Review

A further review of the problems was made by a technical panel.

The members of this panel consisted of members of the Instructional Film Research Program team and representatives from The Ordnance School at Aberdeen Proving Ground.

The purpose of this panel was to review problem scripts again for technical accuracy, review the problems for appropriateness to course objectives, judge levels of difficulty of problems and suggest means for controlling level of difficulty, and to study groups of related items for adequacy of sampling of course content.

When this stage was completed, item scripts were in relatively final form. Shots to be taken by the camera team were detailed on the back of the form. These were arranged to indicate proper editing of the problem. Filming began in March 1952.

6. Review by the Adjutant General's Office Representatives

When approximately 230 such film test problems had been filmed, the Adjutant General's Office requested review of all scripts and of samples of the filmed problems. On the basis of this review the Adjutant General's Office proposed that final selection of problems should depend on the following specific criteria:

a. Problems would be filmed and included if they were "exclusively film" problems, i.e., if they demonstrably embodied crucial movement and/or crucial sound and hence would not be presented in another test medium.

b. Problems would be filmed and included if to present them in another medium would involve complex and ponderous verbal premises.

c. Problems would not be filmed or included if they could reasonably be written as paper and pencil items (including use of still photographs and simple diagrams).

d. Exceptions were approximately 50 problems which could also be presented as paper and pencil items for comparison purposes.

C. Construction of Special Problem Films

1. Power Pack Removal Problem

The aspects of the course most amenable to film testing were the "trouble shooting" (isolation and correction of malfunctions) operations and the procedures involved in performing other operational tests and checks. As the differences between motion picture problems and written

items are analyzed more and more closely, and as the unique characteristics of the motion picture medium are defined, it seems to become quite clear that one of the greatest potentials of the sound motion picture as a test medium is in the area of presenting situations, operations and performances about which subjects can be asked to make discriminating judgments.

The topic selected for the "situation test area" was the removal of the power package from a medium tank. The power pack is a large, heavy and bulky component which requires a crew to efficiently prepare it for removal and to actually remove it from the tank's engine compartment. The problem is set up with a crew of four mechanics at work. These men are plainly identified by large block numbers on their chest and back and small numbers on their jacket cuffs. These numbers (1, 2, 3 and 4) serve as response symbols indicating that the test subject has decided--judged--that some qualitative "thing", which the test subject is looking for, has been performed by mechanic #1 (or 2, or 3, or 4).

The qualities in the work situation, which may or may not be enacted by any one of the four mechanics during a given phase of the work, are:

a. Safety Violation--Did the mechanic do anything that would be likely to result in injury to himself or the other men?

b. Use of Tool--Could the mechanic be given a better tool to work with? Did he use a tool that would result in a waste of time, or in damage to the equipment or in damage to the tool?

c. Unnecessary Job--Did the mechanic do a job that does not have to be done in power pack removal?

The problem is presented in ten steps or stages which serve as time units within which the test subject evaluates the performance of each of the four mechanics on the three qualities described above. Continuous response equipment is desirable but not mandatory in this type of approach to a situational film test. The judgments called for are the same judgments a shop officer, foreman or crew chief would have to make.

2. Field Expediency Problem

Another situational problem was constructed depicting a tank in a mired situation. Specific characteristics of the situation were successively pictured on the screen. The trainee was required to indicate on a check list of 35 activities (Appendix D) the steps he would take to retrieve this particular tank from this particular situation. Both the steps chosen and the order in which they are chosen contribute to the score on this problem.

D. The Criterion

1. Criterion Used

The criterion measure selected for use was the average of the weekly practical grades.

This is a grade obtained weekly by the school from the course instructors for each student. It consists of five graphic type scales: (1) Quality of Work, (2) Application of Classroom Principles, (3) Manual Dexterity, (4) Selection, Use and Care of Equipment, (5) Time (spent completing work). The exact method of combining these scales and of obtaining a quantitative result is not known. For the purposes of the study the weekly score for each student was obtained from the School Registrar and these weekly scores were then averaged.

2. Advantages of Practical Grade Criterion

a. While other possibilities were considered--performance tests, stakes tests, ratings by independent observers, "supervised" instructor nominations, associate ratings and composite weekly grades--each was eliminated as not feasible.

b. The Practical Grade was used instead of the Performance Test Grade because (1) more measures were available of the former (8 to 10 as compared to 3 or 4) and (2) these two instruments were judged to be measuring the same thing in spite of the face validity of the term "Performance Test."

c. Reliability of the criterion seemed quite adequate in comparison with reliabilities reported in other applied studies ($r = .80$).

3. Disadvantage of Criterion Used

a. Restricted Range

A stereotype was plainly evident and rather consistently followed by all instructors that students "just don't fail." The instructors had located the 70 point on the graphic scale and rarely rated a student below it or above 80.

b. Instructor Misconception of Use of Rating

Many of the instructors expressed in one way or another the feeling that "I don't like his attitude so I'll mark him down on the rating" or "He doesn't even try and this is the way I can mark him down." The weekly rating sheets were regarded by the instructors as a fair medium of expressing judgments other than the scales required.

c. Possibility of Contamination

The same instructors who made the grades on floor work also handled and graded the daily assignments and quizzes. (Although the film test was to measure the mechanical proficiency of the student, the criterion nevertheless should be kept free of the influences of results of performance on paper-and-pencil tests.) The weekly written test grade of the student was withheld from the instructors in most cases.

It was common for the instructors to talk about the students among themselves. This could, of course, be a source of contamination but not issuing any futile edicts against it kept the situation more like that in which the test would probably be used.

E. Collection of Data

1. Test Administration Schedule

Graduating classes Nos. 67-81 inclusive that finished the Track Vehicle Repairman Course (9-E-26) at The Ordnance School, Aberdeen Proving Grounds, Maryland, were used as subjects.

The beginning-of-course roster strength figures are given here.

Class Number	N	Class Number	N
67	32	74	22
68	30	75	25
69	30	76	23
70	33	77	25
71	35	78	28
72	27	79	25
73	22	80	38
		81	38

There was anticipated an N of over 400 but due to holdovers, drops, non-use of classes 67-68 and 78, and 12 men who had incomplete papers or scores, a final N of 316 was obtained for the study.

Test administration to Class 67 was in the nature of a "dry run" to iron out administrative problems. Class 68 was lost from the experimental population because of a series of errors in the directions given the troops for reporting locations and times. Classes 69 to 81 inclusive, except for Class 78, were then designated as the experimental population.

Class 78 was not included in the regular schedule. For this class the entire film test was administered in one day in an effort to ascertain the fatigue limits of film testing. The class N was too small to permit statistical conclusions but observations indicated that this amount of film testing (5 1/2 hours running time plus "break time" between reels) is possible in one day.

Two hundred and eight film problems were in the original item pool. Eight of these were eliminated for various technical reasons at different times. The following test presentation schedule resulted: Class 67--208 film problems, Classes 68 to 71 inclusive--205 film problems, Classes 72-81 inclusive--200 film problems (of which three were sample problems). Classes 69-74 inclusive and Classes 75-81 excluding No. 78 were handled as two sub-populations of the total experimental group. The discussion of this is presented in Chapter III.

2. Characteristics of Subjects

The students fell into one of two general categories. Most of them were draftees who had completed basic training and then immediately continued training in this technical course. The second category comprised

men on detached service from an Ordnance Company who would return to their own company for specialist assignment upon completion of the course. Subjectively it can be said that a marked difference in motivation was apparent between the two groups.

a. Age

The subjects ranged in age from 17 to 29 with the mean being 21.9 years and a standard deviation of 2.2 years.

b. Education

Education recorded as highest grade entered, presented a distribution with modes in the two intervals of 8th grade and 12th grade. This is to be expected since these two grades are terminal grades for regular units in our educational system. The mean grade entered was 8.9 with a standard deviation of 2.2 grades.

c. Aptitude Area I Scores

Army Classification Battery Scores comprising Aptitude Area I are interpreted to be representative of the man's ability to learn. One of the sub-tests of the area is Reading and Vocabulary. It is felt that with so much course emphasis and dependence upon the Technical Manuals and verbal-written type of testing that this area is important in relation to course achievement and test responses. The mean Aptitude Area I score of the population tested was 85.94 with a standard deviation of 17.66.

d. Aptitude Area VIII Scores

Army Classification Battery Scores in Aptitude Area VIII are used by the Army as a means of selection of trainees for the Track Vehicle Course. A substantial positive correlation is present between these two area scores (Aptitude Areas I and VIII). A mean score of 91.33 and standard deviation of 17.63 was yielded by the total group tested. As explained above, Classes 67-68 and 78 were excluded from the experimental group and are not included in these figures. This mean of 91.33 indicates that the men enrolled in this course were below the general Army average on Aptitude Area VIII as well as on Aptitude Area I.

c. Motivation for the Film Test

An attempt to provide an incentive for the men to do well on the film test was made by delaying graduation until after the film test. An impression was created that the film test had some effect upon the student's course standing. With this condition we may tentatively assume that motivational conditions relative to passing the test are somewhat equivalent for the film test and paper-and-pencil final tests of the Track Vehicle Course. However, some of the men seemed to be aware that this test would not affect their future careers in the Army.

The differences in attitude between the draftee-pipeline personnel and the RA detached service men were marked. This was observed by the Penn State people as well as the class instructors. Knowledge of assignment, or rumors of it, had a profound effect upon attitudes. The RA men were generally more cooperative and better motivated.

Discipline was not a problem. Only one man was rejected as being unfit to take the test. One man went AWOL during a test session.

Sleeping during the test presented the main problem. Every class would have from 5% to 25% of the class members who would be in varying stages of sleep at different times during the testing period. Most of these soldiers would randomly mark their answer sheets in their moments of semi-awakeness. This problem was met by close and continuous proctoring and eliminating those tests where this type of response was detected as extending over several questions.

III. ANALYSIS AND RESULTS

The purpose of this section is to present in necessary detail the statistical techniques employed in analyzing the data on the motion picture proficiency test and to present the results of the analysis of the data. These techniques dealt with four questions: (1) selection of problems to comprise two comparable forms of a film test which could be used separately, or could be combined into a single test of proficiency; (2) determination of the adequacy of these selected film problems in predicting the criterion measure used; (3) the comparison of the film test and the final examination currently being used at The Ordnance School, which will be referred to as the final school examination in this report, both in terms of their adequacy in predicting the criterion measure and in terms of their relations to other tests; and (4) the comparison of film test problems with paper and pencil items judged to be either (a) comparable to film test problems or (b) to cover comparable material.

A. Problem Selection

Two comparable forms of the test of 59 problems each were derived from an initial administration of 197 problems in the film test pool to 164 students in Classes 69-74. Table 1 presents a summary concerning the degree to which sound and motion were judged to be crucial with respect to the selected 118 problems used in this final film test. Detailed methods used for selecting these problems are given in Appendix L-1.

B. Adequacy of the Film Test

The correlation between the final film test of 118 problems and the criterion used was computed to determine the adequacy of the film test. This correlation was .73 (Table 2). Since the problems for the film test were selected on the basis of statistics computed for Classes 69-74 and the correlation coefficients reported in Table 2 were based on Classes 75-81, the latter may be regarded as cross-validation.

C. Comparison with Final School Examination

To compare the predictive adequacy of the film test and the final examination currently in use at The Ordnance School, a test of the statistical significance of the differences between correlation coefficients was made, using Fisher "z" transformations to take into

TABLE 1

SUMMARY OF SOUND AND MOTION RATINGS OF PROBLEMS ON THE FINAL FILM TEST

Categories	FINAL FILM TEST			FILM TEST I		FILM TEST II	
	No. of Problems	% of Problems in Test	% of Problems[1] in Category	No. of Problems	% of Problems in Test	No. of Problems	% of Problems in Test
4 Motion and/or Sound Crucial	49	42	71	25	42	24	41
3 Motion and/or Sound Significant	20	17	59	10	17	10	17
2 Motion and/or Sound Superfluous	44	37	54	23	39	21	36
1 No Motion or Sound	5	4	42	1	2	4	6
	118			59		59	

[1] This column represents the percentage of problems in each category retained from the original item pool.

TABLE 2

INTERCORRELATIONS, MEANS, VARIANCES AND STANDARD DEVIATIONS OF MEASURES USED IN COMPARISON OF FILM TEST AND ORDNANCE SCHOOL FINAL EXAMINATION (N = 152)

		Final Film Test (118 Problems)	Film Test I	Film Test II	Final School Examination	Practical Rating	Average Weekly Performance Rating	R. V.	A. I.	Composite Rating
		1	2	3	4	5	6	7	8	9
Final Film Test (118 Problems	1	.96*			.77	.73		.48	.73	.84
(59 Problems) Film Test I	2			.92		.71				
(59 Problems) Film Test II	3					.72				
Final School Examination	4				.82 **	.68		.53	.65	.82
Average Weekly Practical Rating	5					.80*	.80			
	$\bar{X}$	63.51	31.23	32.28	66.27	73.90	74.74	83.90	99.52	34.03
	S^2	417.88	111.62	106.04	199.97	10.20	23.47	416.13	441.80	22.83
	S	20.44	10.57	10.30	14.14	3.19	4.84	20.40	21.02	4.78
	$S^2_{\bar{X}}$		.73	.70						

* Corrected by Spearman-Brown Prophecy Formula

C. R. $(z_{15} - z_{45})$ = 1.40 p = .16

C. R. $(z_{19} - z_{49})$ = .75 p = .46

** Computed by Kuder - Richardson Formula 21

C. R. $(z_{47} - z_{17})$ = 1.06 p = .28

C. R. $\bar{x}_3 - \bar{x}_2$ = 3.09 p = .001

consideration the correlation between the correlation coefficients. This test of significance indicated that the difference between correlation coefficients of .68 and .73 was probably not a real difference (Table 2). Correlations of the two tests with the Reading and Vocabulary and Automotive Information Tests of the Army Classification Battery indicated fairly consistent relationships.

D. Comparison of Comparable and Parallel Paper-Pencil and Film Tests

A review of the items included in the regular paper and pencil final examination for this course revealed twenty-three items which were judged to cover the same material as some of the items in the film test. These items and their film counterparts, referred to as comparable tests (Forms A), correlated .61 based on Classes 69-74, and .62 based on Classes 75 to 81 (Class 78 omitted).

An additional fifty paper and pencil test items were constructed to parallel certain of the film items (Forms B). These "sub-tests" correlated .78 and .86 for the class groups mentioned above. The validities of all these sub-tests, presented in Table 10, Appendix L, are fairly comparable and range from .62 to .73.

E. Field Expediency Problem

On the basis of free written responses to the Field Expediency Problem made by students in two classes, a check-list answer sheet was made up. The steps that might be taken in freeing the tank from the mud were listed, and the students were instructed to choose and number the steps they would take. After some experience and "expert" comments, the check-list was revised once.

This check-list in the original or revised form was used with Classes 73-81 inclusive, excluding Class 78. A scoring key for the check-list was developed by personnel of the Ordnance School. Since the students were allowed to choose any number of steps, a differential weighting system was used to distinguish the man who chose five correct steps in five choices, say, from the man who chose five correct steps in fifteen choices. Correct choices were given a weight of two, and incorrect choices were given a weight of minus one.

No attempt was made to "validate" this test. The only results presented in this report are the frequency distribution, mean and variance of the scores for both the original check-list and the revised check-list.

TABLE 3

FREQUENCY DISTRIBUTION, MEANS AND VARIANCES OF RESPONSES TO CHECK-LISTS ON FIELD EXPEDIENCY PROBLEM

Check-List I, Classes 73-75		Revised Check-List, Classes 77-81	
x	f	x	f
-7	1	-7	4
-6	0	-6	2
-5	1	-5	16
-4	5	-4	14
-3	5	-3	24
-2	8	-2	25
-1	10	-1	14
0	13	0	15
1	8	1	9
2	2	2	3
3	4	3	3
4	4	4	2
5	3	5	3
6	1	6	2
7	3	7	0
8	0	8	1
9	0		
10	1		137
	69		
$\bar{x}$ = .2754		$\bar{x}$ = -1.7883	
S^2 = 9.7948		S^2 = 14.7256	

F. Power Pack Removal Problem

The power pack removal film problem was administered to graduating Classes #79, 80 and 81. Since the criterion used for evaluating Film Test I and II represents a measure of performance in the total course rather than for just the aspect of the course covering the power package it was decided that the only data to be presented for this problem would be the frequency distribution, means and variance of the scores. This same data is included for an officer class and a group of instructors to whom the film problem was administered.

The scoring key and answer form used will be found in Appendix E. A student was given 2 for every correct answer and -1 for every wrong answer. Students were instructed to enter a zero (0) in any block where no violation occurred.

G. Student Evaluation of Film Test

It was also thought desirable to obtain a record of the comments made by the students who took the test. In order to have the comments in a form that could be easily summarized, and to reduce the subjective aspect of reporting on verbal comments, a film evaluation form was constructed. This form sought to get students' opinions on those matters which might be important. The answers to these questions were tallied, and the results indicating favorable reactions in general are presented in Appendix J.

TABLE 4

FREQUENCY DISTRIBUTION OF SCORES, MEANS AND VARIANCE FOR CLASSES TESTED ON POWER PACK REMOVAL PROBLEM

Score	Class #79	Class #80	Class #81	Officer Class	Instructor Group	Total Classes 79 - 81
-8 to -5	1					1
-4 to -1	1	1				2
0 to 3	1					1
4 - 7	3			1		3
8 - 11	4	1	2	3		7
12 - 15	2	2	5	2		9
16 - 19	6	1	2	3		9
20 - 23	2		7		1	9
24 - 27	1	3	6	1	1	10
28 - 31		5	6	1	1	11
32 - 35		7	6	1	1	13
36 - 39		6	4	2	4	10
40 - 43		1	1		3	2
44 - 47	1	2			2	3
48 - 51			1			1
N	22	29	40	14	13	91
M	13.3	29.9	26.1	19.8	36.6	24.2
S^2	111.38	112.78	83.64	111.30	49.62	140.26
S						11.8

IV. DISCUSSION

A. Problem Selection

The first of the problems dealt with by the analysis of the data was the selection of items to make two comparable forms of a film test. The item statistics presented in Appendix H need no discussion. The important question in item selection is how good the selected items are when given to a second sample of people. Consequently, the important data in answering the questions about the relative merits of film testing vs. paper and pencil testing are presented in Tables 2, 6, 7, 8, 9 and 10.

The estimated reliability of the film test of .96, and the correlation of this test with the criterion of .73 indicate that the 118 items selected for the film test are highly related to the criterion. However, the frequency distribution of the film test scores indicates that there were possibly too many difficult and easy items, since the distribution appears by inspection to be somewhat platykurtic. Possibly more items of about 50% difficulty should have been included in the film test.

The estimated reliability of .96 for the film test should be interpreted with caution. However, the possibility exists that high reliability may be a characteristic of film tests, because of the concreteness and specificity with which the problems are presented. This specificity may operate to reduce the ambiguity of meaning often present in paper and pencil items. On the basis of the correlation between the two forms (r=.92), it would seem safe to say that the reliability of the film test is high enough to permit prediction for individuals even in cases where a decision made on the basis of the test is a crucial one. The two comparable forms may be reliable enough for individual use, and there is little loss in predictive accuracy when compared with the whole test. Using the Spearman-Brown formula, the estimate of the reliability of thirty items of the sort used in these comparable forms is .85. No empirical check on this estimated reliability has been made, and it seems unlikely that the reliability of such problems as might be used would be this high, if only thirty items were given by themselves.

In summary, on the basis of the evidence presented, the feasibility of sound motion picture tests which yield high reliability has been demonstrated.

From the data on Film Test I and II presented in Table 2, one might conclude that these tests are very nearly comparable. Since Film Test II is slightly easier, one cannot say that the tests are strictly comparable, but for any practical purposes of prediction they may be used as comparable tests. For any experimental work involving comparisons of means of the two tests an adjustment would have to be made, but the two tests seem to be measuring about the same things. The data in Table 1 indicate by inspection that the two forms are comparable based on judgments of crucial sound and/or motion.

B. Adequacy of Film Test and Comparison With Final School Examination

The estimate of the reliability of the final school examination of .82 obtained from these data agrees closely with that of .87 obtained for the same test as reported in Summary of Results 9-E-26 Track Vehicle Repairman's Final Examination for classes for 9 October 1951 to 5 February 1952 by the Training Standards Division of The Ordnance School. The final school examination and either of the comparable film tests take about the same amount of administration time. Since the estimate of the reliability of the final school exam used is an underestimate of the split half reliability of the test, one may conclude that the two methods of testing yield approximately equally reliable tests.

On the basis of the critical ratio of the difference between the correlations of the film test and the final examination with the criterion (1.40) it would seem that the two tests are about equally related to the criterion. When the correlations of the Final Film Test and the Final Examination with the composite score used by The Ordnance School were compared, the two examinations seem to predict this measure equally well. There seems to be little difference between the two tests in terms of prediction of the available proficiency measures. This is further supported by the similarity of the correlation of these two tests with the Reading and Vocabulary Test and the Automotive Information Test.

C. Relation to Reading and Vocabulary Test

One of the original hypotheses was that the film test would minimize the importance of reading in a situation when there were a large number of very poor readers. The statistical test of the correlation of the film test and the final examination with the Reading and Vocabulary Test of the Army Classification Battery did not support this hypothesis (C.R. = 1.06; p = .28).

The hypothesis that the presentation of items through the film medium makes them easier was not adequately tested in view of the fact that order of presentation affects the score to some extent. However, when the film tests were given after the paper and pencil test the mean of the film test was significantly greater (P .001) than the paper and pencil test. On the other hand, when the order of the test was reversed, no significant differences between means were found. (See Appendix L5).

Finally, Table 2 indicates that there is little difference between the two media in terms of prediction of the criterion measure used for this study.

The results of the analysis of the data done for this study indicate that in terms of test construction, it is possible to build a test using sound motion pictures which has quite high reliability and yields a good prediction of the criterion measure that was available for this study. The hypothesis that the film test as constructed would minimize the importance of reading ability and would make items easier to answer relative to a paper and pencil test on the same material was not proved by the data of this study.

D. Field Expediency and Power Pack Removal

The frequency distributions of these tests indicate that they do distinguish between different people, and the tests have considerable face validity. The two tests appear to be methods of testing that should yield useful information, and a further study of their value seems desirable.

V. CONCLUSIONS

1. On the basis of the evidence presented, the feasibility of developing sound motion picture tests which yield high reliability has been demonstrated.

2. There seems to be little difference between the film test and the final paper and pencil examination or between comparable and parallel film and paper-pencil items in the adequacy with which the criterion is predicted.

3. The hypothesis that the film test would minimize the importance of verbal ability was not substantiated.

4. The Field Expediency Test and Power Pack Removal Problem although not validated, do yield score distributions indicating an ability to discriminate among examinees.

5. Film tests are practical to administer, and they can be objectively scored.

VI. SUMMARY AND RECOMMENDATIONS

A. Summary

This is a report of an exploratory study on the feasibility of using sound motion pictures in lieu of conventional proficiency tests. The work reported was done on the Track Vehicle Repairman Course given by The Ordnance School at Aberdeen Proving Ground.

Approximately 200 test problems were constructed after a thorough knowledge of the course had been obtained. The format used was primarily that of the four-alternative multiple choice items. These problems were believed to fall along a continuum from material that may have been tested by verbal items, through materials using still pictures and diagrams, to material that would have been difficult or impossible to test without the use of sound and motion.

The average weekly practical rating of the students by the instructors was the measure with which the results of the film test were compared with the results of paper and pencil tests.

The subjects available for use in selection of problems for the film test and for comparisons of this test with other evaluative instruments were students who completed the course during the period from 10 September to 19 December 1952. These students in general were considerably below the mean of the general Army population as measured by the Aptitude Area I and VIII scores of the Army Classification Battery.

The pool of film problems from which tests were constructed finally included 197 problems. Problems were selected from this pool to make up two comparable film tests of 59 items each.

The reliability of the two comparable tests was .92 as determined by the correlation between them. The correlations of the two tests with the criterion measure were .72 and .71. The correlation of the final examination currently used in the course with the criterion measure was .68, and the estimate of the reliability of this paper and pencil test was .82.

The criterion used for the construction of the final examination was a composite score obtained by differential weighting of weekly paper and pencil quizzes, weekly practical ratings and weekly performance ratings. The correlation of the final paper-pencil examination with this measure was .82; the correlation of the film test of 118 problems with this measure was .84.

The correlation of the film test of 118 problems with the final examination was .77. On the basis of this correlation and the correlation of the two tests with the Reading and Vocabulary and Automotive Information tests of the Army Classification Battery, which were similar for the film test and the final exam, it appears that the film test and the paper and pencil final examination measure much in common.

In order to compare the two media of stimulus presentation when the material tested was the same, 50 of the film problems were given to the students in paper and pencil form, where the stem of the problem was presented verbally rather than pictorially, and the alternative answers were the same. In addition 23 of the items in the final examination were judged to be testing the same material as 23 of the film problems. The difference in medium of presentation did not appear to affect the overall difficulty or validity of these comparable or parallel problems.

Two other kinds of test situations were produced, which seemed to make greater use of the advantage of motion pictures. These problems were attempts to measure behavior on film tests for which no measure of actual behavior was available, and for which paper and pencil tests could not reasonably be constructed.

While the experimental evaluation of these problems was not undertaken they seem to offer great promise as testing procedures, and indicate a kind of situation in which motion picture tests might be used most advantageously.

As a result of the research on which this report is based the following general conclusions may be stated:

1. On the basis of the evidence presented, the feasibility of developing sound motion picture tests which yield high reliability has been demonstrated.

2. There seems to be little difference between the film test and the final paper and pencil examination or between comparable and parallel film and paper-pencil items in the adequacy with which the criterion is predicted.

3. The hypothesis that the film test would minimize the importance of verbal ability was not substantiated.

4. The Field Expediency Test and Power Pack Removal Problem, although not validated, do yield score distributions indicating an ability to discriminate among examinees.

5. Film tests are practical to administer, and they can be objectively scored.

B. Recommendations

This project was an initial attempt to answer a few of a multitude of problems that arise in the development of any new method of testing. It has been carried out in a specific situation in which only a few of the problems of test development, apparent even before the project was started, could be answered. Therefore, it should be looked upon as suggestive of areas in which further work should be done. This section of the report is an attempt to list specific questions for study; some of them were obvious from the beginning while others were defined during work on this project.

1. Areas for Application of Film Tests

The use of motion picture tests could save time and money in situations where extensive performance testing is required. It may well be that the appraisal of performance on relatively small and detailed tasks is not the best situation in which to use film tests. One of the advantages of sound motion pictures is that they can record very complex behavior occurring over a long period of time and in a wide range of places, so that this same complex behavior may be presented repeatedly as a constant stimulus. Thus, the field expediency problem filmed in connection with this project suggests one kind of situation in which film tests should be studied further.

Film testing may be useful in other situations when it is desirable to observe the responses of a large number of people to the same complex situation. Thus, the advantages of film testing may appear in the assessment of personality characteristics where the observation of responses to currently used instruments is confused by constantly changing stimuli, or by the use of stimuli which must be ambiguous in order to be both complex and constant. For example, film tests might be used in the assessment of leadership and supervisory abilities in which a measurement of ability to make complex judgments is considered to be important. Another area for research on motion picture tests would be that of achievement in complex learning situations for people of superior intelligence.

In addition to studying the problem of measurement the relative merits of film tests and paper and pencil tests as material for teaching and review should be investigated. It may be that in many kinds of work, the increased authenticity of a film test used as an aid to teaching could result in greater transfer of learning from the classroom to the point of application of learning. This would imply investigating the effect of immediate knowledge of correct responses to problems, the use of continuous response equipment, and discussion of problems immediately after they have been answered.

Finally, it is suggested that the possibility of using film test problems in conjunction with other types of tests should be investigated.

2. Film Production Practices

It is believed that the feasibility of film tests from the point of view of speed of production and economy of cost depends on using basic film techniques of the kind outlined in Appendix L3. It is therefore recommended that this type of production procedure be followed.

A further economy in production costs could be achieved with some reduction in unessential quality by using typewritten title cards, rather than printed titles. Furthermore film test problems might be constructed at the same time as training films are being produced. Another possibility is to use the same set of visuals with a different sound track to adapt a test problem for a different audience, or a different purpose. It is believed that considerable economy was effected by having an expediting team working ahead of the camera team preparing equipment. It is essential that competent technical advisors be available.

3. Problem Format

In constructing the film proficiency test on the Track Vehicle Repairman Course a number of questions related to test problem format arose which might profitably be pursued. Such questions are:

a. How can the importance of reading and verbal ability in answering a problem be further reduced? It is considered that film tests may be especially appropriate for populations whose reading ability is low.

b. What is the effect of varying amounts of pictorial orientation and sound cues in the problem to be answered?

c. Where should the question stem be located and how often should it be repeated?

d. How should the alternative responses be presented? For example, the alternatives might be repeated by the commentator while the problem is being shown on the screen.

e. How can problems be arranged so that students can make responses while a problem is still on the screen? It is believed that the use of continuous response equipment such as The Classroom Communicator would open up many new possibilities for problem format.

f. Would the repetition of the presentation of a problem before requiring the subjects to respond make the problem easier to solve? Some operations occur rather quickly and a second look might be helpful.

g. What is the effect of varying the length of the response period?

4. Test Administration

Another problem that should be investigated is the advisability of including in the film test complete administration directions in order to insure that these directions are the same for all people taking the test. A comparison should be made of this method with a live test administrator to see if the "personal" touch affects administration.

Study should be undertaken of what appears to be the sleep inducing effects of a film viewing situation. Some of the relevant variables are:

a. Attitude toward viewing technical films

b. Ventilation

c. Amount of light in the room

d. Length of test

e. Attention-gaining devices

f. Daylight projection techniques

g. Manner and technique of presentation of material

5. Further Study of Test Constructed

One problem in connection with the particular test produced for this project is the effect of order of presentation of the problems on response. If a student knows the answer to a question, but gives the wrong answer because of the order in which the question was presented, then his score is in error as an estimate of his knowledge. In this test the problems were presented randomly. The effect of reordering the problems so that they are in homogeneous groups with respect to subject matter should be investigated.

Also in connection with this particular film test, the reliability of the comparable tests should be investigated, when the 59 problems are given by themselves, rather than as part of the pool of 200 problems. These comparable tests should also be further sub-divided, to determine empirically the effect of length of test. It is possible that, particularly with this kind of test, adding additional items might change what an item measures. If the reliability of the film test continues to be high, then film tests should be compared item for item with paper and pencil tests to determine if the film medium yields more accurate scores.

Finally, this film test should be further evaluated using measures of actual field performance.

In summary, all of the problems usually considered in connection with testing should be investigated. It is possible that the use of films as a method of presenting stimuli will yield answers to those problems which are different from those usually obtained because of the difference in media. In addition, the peculiar advantages of motion pictures must be considered in the choice of areas in which motion picture tests could be used profitably to yield better measuring instruments.

REFERENCES

1. Carpenter, C. R., Eggleton, R. C., John, F. T., and Cannon, J. B. The Classroom Communicator. Technical Report SDC 269-7-14, The Instructional Film Research Program, The Pennsylvania State College, 1950.

2. Carpenter, C. R., Eggleton, R. C., John, F. T., and Cannon, J. B. The Film Analyzer, Technical Report SDC 269-7-15, The Instructional Film Research Program, The Pennsylvania State College, 1950.

3. Flanagan, J. C. The effectiveness of short methods for calculating correlation coefficients, Psychol. Bull. 1952, 49, 342-348.

4. Gibson, J. J. Motion Picture Testing and Research. Army Air Forces Aviation Psychology Program Research Report No. 7, Washington, D. C., U. S. Gov't. Printing Office, 1947.

APPENDIX

APPENDIX A

CHRONOLOGY

May, 1951	Invitation to Bid
November	Contract approved and accepted Track Vehicle Repairman Course selected as task
December	Representatives of IFRP studied course Several example film problems shot
January, 1952	Analysis of course continued Preliminary draft completed of approximately 400 problems Review of problem scripts by IFRP Staff Investigation of performance tasks as criterion
February	Review of problem scripts by senior instructors, TOS Technical review of problem scripts by administrative personnel, TOS Course Outline developed Requests for practical performance testing as criterion
March	Photography of problems begun
April	Photography continued Conference at Penn State with representatives of TAGO which resulted in change of interpretation of contract objectives
May	Power Pack Removal Problem photographed Photography completed at Aberdeen Sound recording begun Rater training begun Rough-cut editing of problems completed
June	Sound recording completed at Aberdeen Rater training completed Sample problems pre-tested on students Editing continued First experimental class begun course instruction Collection of criteria data begun Review of problems by representative of TOS and TAGO at State College Request for team of raters to collect criterion data. The request could not be granted.

July	Narration recorded Sound recording and laboratory processing completed
August	Composite prints of main problem pool completed Parallel and comparable paper-pencil items prepared Editing of Field Evacuation Problem completed Editing of Power Pack Removal Problem continued
September	Testing begun Scoring begun Editing of Power Pack Removal Problem completed
October	Testing and scoring continued
November	Testing and scoring continued Statistical analysis begun
December	Testing and scoring completed Statistical analysis continued Writing of final report begun
January, 1953	Statistical analysis completed First draft of final report submitted Two prints and regative of all film test problems delivered to TAGO
February	Revised full final report submitted
March	Revised condensed report submitted and approved
April	One print of all test problems delivered to The Ordnance School, Aberdeen Proving Ground Printed final report delivered to TAGO

APPENDIX B

Form IFRP - AGO - 2

AGO Film Test Project

TANK MAINTENANCE TRAINING COURSE
OBSERVATIONAL REPORT OUTLINE

Operation _____ Report Serial Number _____ Item No. _____

Purpose of Item _____

_____

_____

Statement of Item Content _____

_____

_____

Premise _____

_____

Correct Answer _____

_____

Distractors _____

_____

_____

_____

Item Writer _____ Date _____ Reviewer _____ Date _____

Expert _____ Date _____ Expert _____ Date _____

Photographed _____ Date _____ Shooting Time _____

Film Review _____ Date _____

APPENDIX C

COURSE COVERAGE BY SUBJECT AREA

A course outline was developed based upon the expert opinion of those responsible for the course of the relative importance of the many topics covered in the course.

This course outline is condensed and simplified in Table 5. In this table the relative weights to be given to each subject area for the ideal course coverage are expressed in percents. These may be compared to the actual weight given the areas in the pool of two hundred film problems, and the two comparable tests of fifty-nine problems each which emerge from that pool. It will be seen that the course has been reasonably well covered by the film tests.

An item frequently tested more than one subject area and was therefore classified according to the areas primarily and secondarily sampled. For instance, a problem primarily about engines might, secondarily, require the student to have some knowledge of the electrical system.

TABLE 5

COURSE COVERAGE BY SUBJECT AREA

Subject Area	Req. % For Coverage	Total Test (200 Items) Prim. Class.	Total Test (200 Items) Sec. Class.	Film Test I Prim. Class.	Film Test I Sec. Class.	Film Test II Prim. Class.	Film Test II Sec. Class.
Field Expediency	11	6	3	5	3	5	2
Basic	10	11	6	10	5	10	7
Engines	15	16	13	15	14	15	12
Fuel System	7	6	5	9	6	9	7
Elect. System	14	16	25	17	22	17	22
Power Train	24	24	20	25	22	24	16
Susp. System	8	5	5	5	5	6	8
Hull Comp.	3	6	6	4	5	5	7
Malfunction	3	10	16	10	18	9	18
Preventive Main.	4	0	1	0	0	0	1
Interchangeability .	1	0	0	0	0	0	0
	100	100	100	100	100	100	100

APPENDIX D

SPECIAL ANSWER SHEET FOR FIELD EXPEDIENCY PROBLEM

Name	Rank	Serial Number

AGO FILM PROFICIENCY TEST

Special Answer Sheet Number 1

You are going to see a movie of a tank that is stuck in the mud. Your problem is to get the tank out as quickly as possible, using only the things you can see in the pictures.

To get the tank out, you would have to follow in the right order some of the steps listed on the next page. Your problem is: What are the steps? In what order should the steps be taken?

Read over the list on the next page and indicate your choice of steps by numbering each step you would take. Number the first step you would take "1"; number the second step you would take "2" and so on. The steps from which you are to choose are listed on the next page; if you wish to use some steps not listed, write them on the back of the second sheet, and assign them numbers according to the point at which you would take them.

You may turn to the next page when you are sure you understand what you are to do.

_____Cut down some trees

_____Drop a tree behind two trees that are beside each other

_____Make a deadman

_____Attach a cable to a standing tree

_____Dig out around tracks

_____Get barbed wire from old fence

_____Attach cable to pintle

_____Attach cable to right track

_____Attach cable to left track

_____Attach cable to log of deadman

_____Attach cable to both tracks

_____Attach cable to sprocket drive wheels

_____Put rocks under both tracks

_____Put rocks under left track

_____Put rocks under right track

_____Put brush under both tracks

_____Put brush under left track

_____Put brush under right track

_____Put logs under right track

_____Put logs under left track

_____Lock left track

_____Lock right track

_____Block in rear of tank to avoid slipping backwards

_____Tie one log across both tracks at front with cable

_____Tie one log across both tracks at front with barbed wire

_____Tie logs across both tracks at front with cable

_____Tie logs across both tracks at front with barbed wire

_____Tie one log across both tracks at rear with cable

_____Tie one log across both tracks at rear with barbed wire

_____Tie logs across both tracks at rear with cable

_____Tie logs across both tracks at rear with barbed wire

_____Move tank forward at low speed

_____Move tank backward at low speed

_____Move tank backward at high speed

_____Move tank forward at high speed

_____Repeat steps numbered_____

APPENDIX E

SPECIAL ANSWER SHEET FOR POWER PACK REMOVAL PROBLEM

The Power Pack Removal Problem was answered by the test subjects on a prepared answer sheet. A copy of this answer sheet, reduced in size, may be found on the following page. The test subjects were instructed to enter their responses for each step of the problem in the appropriately numbered line of the answer sheet. Recording the number (1, 2, 3 or 4) of one of the mechanics shown at work in the problem in one of the answer blocks indicates that that mechanic committed the error identified in the column heading for the step indicated on the left. For example, a 3 entered in the block under column C of line five indicates that Mechanic #3 did an Unnecessary Job in Step 5 of the problem. A zero (0) means that none of the mechanics committed a violation. The correct responses are filled in on the answer sheet.

Due to lack of time and absence of criterion measure no intensive study was made of possible scoring procedures. The method outlined above is only a suggestion. If this problem is used in the future it is urged that prior investigation be made of scoring procedure. On ground of rationale, for example, it is suggested that a scoring system be used as follows: (1) the answer sheet be similar in form to the one used, (2) the numbers 1, 2, 3 and 4 appear already entered in each block, (3) for each observed violation in any specific step the number of the mechanic committing the violation would be crossed out (or encircled). This scoring procedure would allow inferences to be drawn as to the intent of the test subject where no response is made in a particular block. It also weights each block equally since a pattern of four responses is established for each block as contrasted with the scoring procedure used above.

	A SAFETY VIOLATION	B BETTER TOOL	C UNNECESSARY JOB
Step 1	S A M P L	E S	T E P
Step 2	S A M P L	E S	T E P
Step 3	3, 4	4	1, 2
Step 4	1, 2	4	0
Step 5	0	0	3
Step 6	1	0	0
Step 7	2	0	0
Step 8	1, 2	0	0
Step 9	1	0	0
Step 10	1, 2, 3, 4	0	0

APPENDIX F

COMPARABLE PAPER-PENCIL TEST - FORM B

Track Vehicle Repairman Course 9-E-26

1. What does the impulse coupling do on the magneto of the Wisconsin auxiliary engine?

    a. Limits the magneto voltage output
    b. Improves the spark at low rpm
    c. Insures a constant point gap
    d. Acts as a governor

2. What would probably cause foaming oil to flow from the oil breather of the M26 tank?

    a. Front oil pump sucking air
    b. Defective regulator valve
    c. Dirty oil in transmission
    d. Front oil pump producing too much pressure

3. What should be done next when valve timing the AV 1790 engine, if the timing mark on the flywheel is in line with the pointer, and the cam roller is free to turn?

    a. Set intake valve at .100 clearance
    b. Turn flywheel counterclockwise 1/8 turn
    c. Install cam drive quill
    d. Align magneto mark with pointer

4. What would a mechanic be doing wrong when using an Allen wrench to remove the oil pressure check valve plug on the cross drive transmission with the engine running at low rpm?

    a. Has the engine running
    b. Has the engine rpm too low
    c. Removing a plug that should never be removed
    d. Using the wrong type wrench

5. What should be adjusted when the M 26 steering brake control handle is measured and there is excessive travel?

   a. Clutch linkages
   b. Parking brake locks
   c. Brake clevis yokes
   d. Differential bands

6. For which of the following purposes is an "easy out" tool used?

   a. Removing a Rosan insert
   b. Enlarging and tapping cylinder head threads
   c. Cleaning out cap screw threads
   d. Removing a bronze bushing

7. What is wrong with an engine which gives a momentary break in performance if the gas pedal is suddenly depressed when the engine is idling?

   a. Worn fuel pump
   b. Faulty choke
   c. Defective accelerating pump
   d. Blocked idling jet

8. If a radial engine has piston slap, which of the following would be the most likely cause?

   a. It is cold
   b. It is too hot
   c. Sticking valves
   d. The tappets need adjustment

9. Why would a mechanic hand crank a radial engine before turning on the ignition and pressing the starter switch?

   a. To prime the cylinders
   b. To wind up the magnetos
   c. To check for hydrostatic lock
   d. To circulate the oil

10. What is probably wrong with the M 24 if the starter won't crank the engines? The lights are on, but dim too much when the strater is pressed.

    a. The battery ground circuit is open
    b. The starting circuit is open
    c. The ignition switch is open
    d. The battery is run down

11. What should be done in adjusting transfer case linkage after removing the shifter-shaft extension clamp bolt?

    a. Loosen lock nut on shift rod
    b. Remove cotter pin and washer
    c. Center the shift lever
    d. Free the shifter-shaft extension

12. Which of the following is a proper way to test for a broken torsion bar?

    a. Use a jack under the road wheel
    b. Pull up on the road wheel
    c. Check the road wheel support arm nut with a wrench
    d. Use a pry bar under the road wheel

13. Which of the following best describes the principle of magnetism?

    a. Attraction of like poles
    b. Attraction of opposite poles
    c. Adjacent magnetic repulsion
    d. Mutual magnetic repulsion

14. Why must the release levers be adjusted and checked with a ruler when assembling the Ford clutch?

    a. To make sure that there is a uniform pressure on the plate
    b. To make sure the levers are not worn excessively
    c. To make sure all of the levers are free to move in the same direction
    d. To make sure the clutch levers do not bind

15. What should be done next in adjusting the hydra-matic transmission if the rear servo gage short end is hooked over the spring end of the servo and there is a gap between the long end of the gage and the rear servo piston rod?

    a. Tighten down the lock nut with special wrench
    b. Turn band adjustment indicator rod cap
    c. Relieve spring tension using a screwdriver
    d. Loosen adjustment screws evenly

16. What is probably the trouble with the M 24 tank when the starter is pressed, if the lights go out, and the starter doesn't crank the engine?

    a. Open switch in starting circuit
    b. Open battery ground circuit
    c. No ground in starter relay circuit
    d. Loose battery connections

17. Into what range should the transmission lever of the M 24 tank be placed if the transfer shift lever is in reverse?

    a. Low
    b. Drive
    c. Neutral
    d. Reverse

18. What part of the electrical system is checked using the high voltage circuit tester?

    a. Starter circuit
    b. Battery-generator circuit
    c. Voltage regulator circuit
    d. Secondary circuit

19. What is probably wrong with a rubber tank track that loses springiness?

    a. Sprocket is out of line
    b. Track blocks are uneven in weight
    c. Center guide nuts are loose
    d. Grommet on pin is worn

20. With the M 24 engine running, what does oil bubbling out of the transfer unit indicate when the pipe plugs are removed?

a. Oil pressure is probably too low
b. Transfer unit pump is satisfactory
c. Transmission gaskets are bad
d. Oil flow is reversed

21. What happens to the planet carrier if the M 46 is moving at constant speed and the steering shaft of the cross drive transmission is stopped?

a. Speed up
b. Stop
c. Slow down
d. Its speed will not change

22. What would be indicated if a spark plug on the Cadillac engine was removed and fluid poured out of the cylinder?

a. Sticking valve
b. Hydrostatic lock
c. Worn rings
d. Blocked oil line

23. What should be checked first when the M 24 engine will not start if the lights aren't dimmed by the starter, and the transmission warning lights are out?

a. Master relay circuit
b. Transmission oil signal switch cable
c. Battery master switch
d. Battery ground circuit

24. What will happen if the package unit of the cross drive transmission is installed with the jumper tubes missing?

a. Steering drive gear will bind
b. Hydraulic system will not function properly
c. The vehicle will operate normally
d. Locating dowel cannot be aligned

25. What is indicated when a cylinder gives the same low reading for the "wet" compression test?

    a. Leaking valves
    b. Worn piston rings
    c. Cracked piston
    d. Scored cylinder walls

26. What causes excessive arcing at the circuit breaker points of a distributor?

    a. Faulty rotor
    b. Faulty cam
    c. Faulty condenser
    d. Faulty coil

27. What could be the trouble when the M 24 throttle pedal is pressed down hard and no change occurs in engine noise or tachometer?

    a. Jammed degasser
    b. Blocked air cleaner
    c. Blocked high speed circuit
    d. Float level too high

28. What is a mechanic checking by removing the warning signal switch of the M 24 and installing a pressure gauge?

    a. Main line pressure
    b. Lubrication pressure
    c. Converter pressure
    d. Low range pressure

29. What is wrong if the tank engine is running at 1000 rpm, no lights are on, the ammeter reads zero, and a test of the generator shows a 29 volt output?

    a. Defective ammeter
    b. Reversed polarity of generator
    c. Discharged batteries
    d. Loose generator drive belts

30. What is probably wrong with the M 24 when the starter is pressed but does not crank, the lights do not dim, and the transmission warning lights are on?

    a. Open starting circuit
    b. Starter pinion not engaging
    c. Battery ground circuit open
    d. Battery run down

31. Which of the following is a right way to do a job on the Ford engine?

    a. Clean the cuno filter with a brush
    b. Replace the push rods in random order
    c. Remove the starter with engine upside down
    d. Tape the connecting rod bolts

32. What should be adjusted if a low voltmeter reading is found when checking the M 26 voltage regulator?

    a. Reverse current cut-out relay
    b. Silverstat fingers
    c. Voltage rheostat
    d. Ballast lamp

33. Why would a mechanic use a helicoil on a AV 1790 engine?

    a. Provide sturdier grip in soft metal
    b. Clean out spark plug threads
    c. Adapt holes for different size plugs
    d. Replace Rosan inserts

34. What might be shown if there is no charging rate on the M 24 ammeter, but a reading occurs when the battery and armature terminals of the regulator box are connected?

    a. Generator armature defective
    b. Circuit breaker relay defective
    c. Loose connection in charging circuit
    d. Burned out field windings

35. What should be checked first if the T 41 tank engine stopped running when you switched from one magneto to the other?

a. Booster coil
b. Condenser
c. Spark plug
d. Magneto

36. Why would a mechanic have trouble installing a torsion bar with the alignment mark other than straight up?

a. The bar wouldn't go in that place
b. The bar must have been put in wrong end first
c. It should have been screwed in direction of the arrow
d. The blind spline would not line up

37. Where is the trouble if the LVT engine does not cut off until the control panel cover is removed and the magneto switch is grounded?

a. In the high tension wiring
b. In the booster circuit
c. In the magneto itself
d. In the magneto switch

38. What would cause the LVT radial engine to misfire regularly?

a. Blocked fuel filter
b. Defective spark plug
c. Faulty ignition timing
d. Defective fuel pump

39. What magnetic principle is shown by a closed coil circuit with a small lamp attached which lights up as it passes near a strong magnetic field?

a. Mutual induction
b. Self-induction
c. Adjacent induction
d. Eddy currents

40. What would a mechanic be doing wrong by cleaning the M46 tank engine with a brush and paint thinner?

a. Nothing, he'd be doing the job right
b. He would be using the wrong material to clean the engine
c. He should use a spray instead of a brush
d. Nothing, if he keeps the thinner off the wiring harness

41. What is likely to happen if the Ford camshaft nuts are loosened and completely removed from left to right?

    a. Push rods will be damaged
    b. Bearing cap will break
    c. Threads on bolts will strip
    d. Camshaft may be bent

42. Which of the following is the proper way to stop the M46 tank engine?

    a. Cut off fuel, turn magnetos off
    b. Turn magnetos off, cut off fuel, turn off master switch
    c. Turn off magnetos, turn off master switch
    d. Cut off fuel, turn magnetos off, turn master switch off

43. Why would a front servo gauge be used on the hydra-matic transmission?

    a. To open front band for adjustment
    b. To apply pressure for band adjustment
    c. To force front clutch into engagement
    d. To equalize clutch pressure

44. What should the mechanic do next on the controlled differential after turning the steering brake band adjusting nut counterclockwise 2 half turns?

    a. Run a stall test on engine
    b. Check pull back tension of steering brake
    c. Check transfer shift linkage for binding
    d. Replace lock nut over adjusting nut

45. What, if anything, would a mechanic do wrong by hammering the brake-apply-cam stationary ring counterclockwise on the end cover of the cross drive transmission?

    a. Nothing, he'd be doing the job correctly
    b. He should make sure the sun gear is removed first
    c. He should be rotating the ring clockwise
    d. He should not use a hammer

46. What would a mechanic be doing when he flashes the field terminal of the M26 generator?

a. Making battery flash test
b. Polarizing the generator
c. Checking battery condition
d. Checking output of generator

47. What is indicated if the hydraulic valve lifter plunger is depressed by finger but does not bounce up until tapped on the bottom?

a. Check - ball is worn
b. Piston is worn
c. Cylinder is worn
d. Leak down is improper

48. What should the mechanic do next after lowering the end cover of the cross drive transmission to the table for disassembly?

a. Remove output flange
b. Lift out the planet carrier
c. Level the end cover
d. Remove brake adjusting bracket

49. Why is the M32 lifting cable attached to the side of the hull after the boom is raised?

a. To prevent boom falling if cable breaks
b. To strengthen boom under load
c. To stow lifting cable during recovery
d. To provide boom stabilizer is load shifts

50. What should be done next when checking hydrostatic lock on the Cadillac engine after the gasoline has been drained from the cylinders and the engine has been hand cranked?

a. Remove and drain muffler
b. Check crank case oil level
c. Start other engine to ventilate engine compartment
d. Start engine with the starter

OS Dir Tng 36 - 4 Sep 45

PRACTICAL GRADE

CLASS NO. NAME

QUALITY OF WORK (Accuracy, precision of actual operations performed)	Much below quality requirements	A little below quality requirements	Just meets quality requirements	A little better than quality requirements	Much better than quality requirements
APPLICATION OF CLASSROOM PRINCIPLES (Ability to apply to his work what he has been taught in class)	Makes no use of classroom principles	Uses some of the classroom principles	Applies the average amount of classroom principles	Makes use of nearly all classroom principles	Applies all classroom principles
MANUAL DEXTERITY (Skill in use of the hands)	Very poor use of hands. Clumsy. Can't coordinate mind and hands	Below average ability to use hands	Has average amount of ability to use hands	Above average in the use of hands	Excellent use of hands. Very agile. Excellent coordination of hands and mind.
SELECTION, USE & CARE OF EQUIPMENT (Safety, cleanliness, proper adjustment and control of equipment)	Continually misplaces equipment. Incorrect selection or adjustments. Dangerous worker.	Usually spoils equipment or makes poor selection & adjustments. May take careless chances.	Just meets requirements of safety selection & adjustment of equipment. Seldom takes chances.	Usually careful about equipment. Selects & adjusts well. Takes few chances.	Conserves materials. Always careful about equipment. Excellent selection & adj. Takes no chances.
TIME (Relative time spent completing his operations)	Much behind time requirements. Slow worker.	A little behind time requirements.	Just meets time requirements. Average speed.	Finishes a little sooner than required.	Finishes sooner than time requirements

Quality of work _____

Application of classroom principles _____

Manual Dexterity _____

Selection, Use and Care of Equipment _____

Time _____

TOTAL _____

_____

Signature of Instructor

APPENDIX G-1
Practical Grade Sheet

ARMY OD ABERDEEN PROVING GROUND MD 431

ORDOS-1256--12 Sept 51
(Formerly OS Tmo 371

PERFORMANCE TEST GRADE CLASS NO. NAME

QUALITY OF WORK (Accuracy, precision of actual operations performed)	Much below quality requirements.	A little below quality requirements.	Just meets quality requirements.	A little better than quality requirements.	Much better than quality requirements.
PLANNING (Problem solving ability, analytical or systematic attack)	Solves no problems. No planning.	Solves only an occasional easy problem. Some planning.	Just meets requirements of problem solving and planning.	Solves nearly all problems. Works systematically.	Solves all problems. Excellent planning.
TIME (Relative time spent completing the operation)	Much behind time requirements.	A little behind time requirements.	Just meets time requirements.	Finishes a little sooner than required.	Finishes sooner than required.
USE & CARE OF TOOLS (Selection, use and care of proper tools, references and supplies. Safety)	Continually misplaces or breaks equipment. Incorrect selections. Dangererous worker.	Usually spoils equipment or makes poor selection. May take careless chances.	Just meets requirements of use and care of equipment. Selects fairly well. Seldom takes chances.	Usually careful about equipment. Selects well. Takes few chances.	Conserves materials. Always careful about equipment. Excellent selection. Takes no chances.

Quality of work - _____
Planning - _____
Time - _____
Use and care of Tools - _____
TOTAL - _____

_____
Signature of Instructor

ARMY---OS---ABERDEEN PROVING GROUND, MD---933

APPENDIX H

ITEM STATISTICS FOR 200 FILM TEST ITEMS

This appendix includes the statistics for the 197 problems which were used as the problem pool from which the final film test was constructed. Problems 1, 2, and 3 in the film problems pool were sample problems to which the answers were given by the narrator, so they are omitted from this Appendix. The problem statistics are based on Classes 69 - 74 inclusive (N = 164).

The problem numbers are given in the first column. Some of the problems were moved from their original location in the film; the number given first is the problem number in the film as it now is, the number in parentheses being the number in the original film. Those problems chosen for the final film test are indicated in column 2. Those problems which are marked with a Roman numeral I are in Film Test I; those marked with a Roman numeral II are in Film Test II. Where there is no number in column two, the problem was not part of the final film test.

Column 3 gives the difficulty of the problem; this is the proportion of the total group of 164 subjects who answered the problem correctly. In column 4 are listed the estimates of the biserial correlations of the problem with the criterion measure. The method of computation is described in Appendix L-1.

The next four columns give the frequency of response for each of the alternative answers, with the correct answer indicated by a star. The last column gives the rating of the problem on a four point scale with respect to sound and motion, four indicating that sound and motion is crucial. The way these ratings were made, and the exact meaning of each rating is described in Appendix L-1.

APPENDIX H

ITEM STATISTICS FOR 200 FILM TEST ITEMS

Item No.	Film Test No.	p	r_{bis}	Frequency of Response for Alternative Answers				Rating on Sound and/or Motion
				1	2	3	4	
4	II	.71	.33	116*	14	16	15	3
5	I	.43	.45	40	17	34	70*	4
6	II	.35	.36	24	57*	25	56	4
7		.32	-.17	49	53*	38	21	2
8		.44	.10	7	77	72*	6	2
9	I	.65	.30	5	39	106*	9	3
10		.28	.08	46*	62	27	26	2
11	I	.72	.33	2	17	24	118*	3
12	II	.34	.39	56*	50	22	34	4
13		.33	.03	18	54*	73	18	3
14	I	.57	.42	93*	5	56	6	4
15	I	.75	.25	5	8	26	123*	3
16		.42	.07	44	2	46	69*	2
17		.10	.10	11	17*	47	81	4
18	II	.88	.42	5	3	9	145*	2
19	I	.88	.35	3	0	144*	14	4
20		.54	.24	88*	61	4	8	2
21		.32	.04	31	40	52*	38	2
22	II	.41	.38	7	59	28	68*	2
23	I	.54	.51	9	20	89*	44	4
24	II	.56	.36	26	32	13	92*	4
25	I	.85	.43	140*	9	7	8	4
26	II	.65	.36	10	107*	9	37	2
27	I	.71	.60	15	11	20	116*	4
28		.38	.09	24	41	62*	34	3
29		.58	.22	17	13	95*	38	4
30		.14	-.26	23*	26	87	25	2
31		.37	.24	61*	24	41	37	4
32		.41	.24	62	18	13	67*	2
33		.58	.20	15	9	95*	42	3
34	I	.45	.56	49	73*	21	18	2
35		.35	.25	56	58*	36	11	4
36	II	.30	.41	49*	70	28	14	2
37	I	.44	.42	27	72*	35	27	4
38	I	.68	.29	22	21	112*	7	2
39	I	.34	.22	15	12	56*	79	4
40		.28	-.05	17	18	46*	80	2

Item No.	Film Test No.	p	r_{bis}	Frequency of Response for Alternative Answers				Rating on Sound and/or Motion
				1	2	3	4	
41		.47	.05	9	61	77*	13	3
42	II	.60	.35	2	38	33	99*	2
43		.57	.18	37	93*	25	7	1
44		.62	.28	17	101*	21	22	1
45	II	.40	.33	38	14	35	66*	1
46	II	.16	.35	13	26*	80	42	4
47	II	.62	.45	12	9	37	101*	2
48		.24	-.11	17	45	55	40*	4
49		.30	.31	65	17	49*	6	1
50		.40	.14	44	22	65*	26	3
51	I	.32	.31	53*	5	69	30	3
52	I	.54	.44	30	22	88*	18	4
53	II	.33	.45	7	21	54*	76	2
54	I	.51	.57	9	11	83*	54	4
55	I	.29	.53	64	29	48*	16	2
56		.41	.20	13	14	66	67*	4
57		.17	.06	45	28*	77	22	2
58		.42	.15	19	69*	33	36	4
59	I	.56	.29	92*	14	38	15	3
60	II	.59	.29	23	97*	19	20	2
61	I	.73	.56	119*	22	15	4	2
62		.43	.20	71*	49	16	22	2
63	I	.60	.69	99*	3	15	43	2
64	I	.56	.49	7	7	92*	54	2
65	I	.68	.46	6	12	29	111*	2
66	I	.56	.54	42	11	91*	27	2
67	I	.22	.49	36*	34	7	87	1
68		.63	.19	103*	21	18	22	2
69	I	.48	.45	78*	23	26	36	4
70	I	.58	.54	23	95*	24	19	2
71	II	.62	.33	31	102*	16	17	2
72		.27	.08	16	44	59	45*	2
73		.24	.13	39*	5	62	56	3
74		.37	.25	72	60*	21	11	3
75	II	.60	.24	16	98*	17	31	2
76	II	.58	.52	13	95*	37	17	1
77		.45	.14	73*	61	17	11	4
78	I	.68	.47	10	112*	21	19	2
79		.73	.10	10	120*	21	11	2
80	I	.59	.48	96*	16	22	28	2

Item No.	Film Test No.	p	r_{bis}	Frequency of Response for Alternative Answers				Rating on Sound and/or Motion
				1	2	3	4	
81		.25	.08	55	39	32	41*	2
82	II	.48	.29	16	30	38	78*	2
83	II	.76	.36	3	10	25	124*	2
84		.26	.20	59	45	42*	17	2
85		.63	.20	8	33	19	103*	2
86		.78	.12	17	9	8	128*	4
87		.38	.29	30	37	63*	31	2
88	II	.59	.51	97*	8	22	35	2
89	II	.54	.48	24	89*	26	19	4
90(205)	I	.57	.35	11	13	46	93*	4
91		.15	-.01	24*	17	86	32	2
92(206)	I	.66	.43	12	7	35	109*	4
93		.51	-.28	24	14	84*	58	3
94	II	.51	.34	83*	24	6	47	4
95		.30	-.05	21	28	64	49*	2
96		.42	.17	19	34	41	69*	2
97	II	.68	.29	35	7	112*	10	1
98		.11	.19	18*	33	44	68	4
99	II	.90	.34	8	4	2	148*	3
100		.45	.13	27	74*	58	4	2
101	I	.61	.44	3	14	46	100*	4
102	II	.68	.58	111*	13	20	19	4
103	I	.48	.37	26	78*	13	46	3
104	II	.80	.46	3	13	16	131*	2
105(208)	II	.71	.52	117*	10	29	6	4
106		.29	.28	53	58	47*	6	2
107	I	.58	.31	9	9	95*	54	2
108		.24	.29	53	53	40*	16	2
109		.26	.35	42*	22	82	17	2
110		.53	-.16	87*	18	28	30	1
111		.27	-.11	41	45*	65	14	4
112		.24	.10	14	58	40*	51	2
113	I	.58	.43	95*	14	24	30	3
114	II	.65	.48	9	106*	30	16	2
115		.30	.35	35	50*	29	53	1
116	II	.64	.48	14	105*	17	27	3
117	I	.80	.53	6	16	11	131*	3
118	II	.58	.59	14	95*	24	30	4
119		.29	.21	48*	53	45	18	4
120		.40	.19	6	65*	27	67	2

Item No.	Film Test No.	p	r_{bis}	Frequency of Response for Alternative Answers				Rating on Sound and/or Motion
				1	2	3	4	
121	II	.45	.29	5	73*	11	74	4
122	I	.49	.27	18	22	81*	42	2
123	II	.68	.30	13	112*	13	25	4
124	II	.64	.49	105*	12	10	36	4
125	I	.46	.35	76*	21	29	40	2
126	I	.33	.61	54*	8	22	79	4
127	II	.77	.41	16	12	127*	10	3
128	II	.45	.26	73*	24	13	53	2
129		.59	.21	16	46	96*	8	2
130	I	.52	.48	30	85*	18	30	2
131	II	.56	.56	92*	10	36	25	2
132	I	.38	.35	14	63	23	63*	2
133	II	.58	.50	95*	12	27	31	1
134		.05	-.04	24	8*	114	19	4
135	II	.49	.51	80*	28	26	29	2
136		.25	.11	41*	56	9	58	4
137	II	.57	.42	94*	25	14	32	4
138	II	.35	.26	14	57*	37	55	4
139	I	.69	.53	21	24	5	113*	4
140	I	.61	.48	32	13	100*	17	4
141	I	.27	.39	34	57	45*	26	4
142	II	.58	.42	12	95*	24	32	4
143		.10	.05	20	69	17*	54	2
144	I	.74	.56	20	122*	13	8	3
145	II	.45	.52	73*	12	16	61	4
146	II	.77	.52	11	9	127*	16	2
147	II	.41	.38	68*	8	17	68	3
148	II	.71	.46	7	117*	28	8	3
149		.08	-.41	13*	60	23	65	4
150	II	.60	.60	98*	33	17	15	2
151	I	.23	.37	1	38*	107	14	2
152		.10	-.38	17*	81	3	60	4
153		.36	.01	59*	26	57	19	3
154		.38	.17	61	63*	31	9	2
155	I	.37	.38	4	61*	87	11	4
156		.15	.10	25*	7	4	125	4
157	I	.58	.62	30	95*	17	21	2
158	I	.63	.52	104*	25	18	14	2
159	II	.74	.35	6	12	22	122*	4
160		.79	.09	129*	9	16	7	2

Item No.	Film Test No.	p	r_{bis}	Frequency of Responses for Alternative Answers				Rating on Sound and/or Motion
				1	2	3	4	
161		.37	.05	2	22	61*	77	3
162		.74	-.08	28	121*	3	8	4
163	II	.68	.53	112*	15	26	10	3
164		.39	.02	9	64*	62	28	2
165	II	.68	.42	20	6	24	112*	3
166		.57	.23	23	94*	24	22	2
167	II	.52	.46	12	86*	44	20	4
168		.18	.37	59	46	29	29*	2
169	I	.63	.39	15	9	103*	35	4
170	I	.70	.37	2	17	114*	29	2
171		.69	.22	12	22	16	113*	2
172		.69	.22	8	39	113*	3	3
173	II	.65	.38	106*	20	28	11	4
174		.50	.22	82*	32	30	16	2
175(202)		.54	.22	30	11	32	89*	3
176	II	.45	.23	4	73*	43	36	4
177	I	.27	.40	20	46	44*	52	3
178		.30	-.10	37	50*	40	32	1
179(204)		.38	.28	37	62*	22	43	2
180	II	.55	.45	12	90*	46	12	2
181(201)	II	.51	.32	48	18	13	84*	4
182		.49	.08	20	53	10	80*	4
183	II	.30	.40	49*	30	20	63	4
184	I	.45	.44	43	74*	27	19	4
185		.29	.21	84	48*	3	29	2
186	II	.26	.30	43*	56	33	31	4
187		.37	.06	24	38	60*	41	4
188	II	.45	.51	23	27	39	73*	3
189(203)	II	.40	.34	36	48	65*	14	3
190	I	.56	.30	13	91*	33	28	4
191		.44	.20	9	70	13	72*	1
192		.52	.32	85*	60	13	2	3
193	I	.23	.29	34	37*	24	69	4
194		.48	.06	79*	49	18	18	3
195	I	.79	.41	8	4	21	129*	2
196	II	.68	.54	18	18	15	112*	4
197	I	.71	.33	3	20	21	117*	2
198	I	.45	.28	30	73*	23	35	2
199	I	.59	.37	16	30	21	96*	4
200(207)	I	.45	.29	31	74*	17	40	4

APPENDIX J

FILM TEST EVALUATION FORM

The film test evaluation form completed by the subjects in Classes 71 - 81 inclusive excluding 78 is presented in this Appendix, with a summary of the answers obtained. The figures given represent the proportion of students answering a particular question who chose each answer.

Questions 5, 9, and 12 have sub-parts, to be answered only if a particular response was given to the main question. The percentages given for their sub-parts represent the proportion of students answering the sub-question who made a particular response to it.

APPENDIX J

AGO FILM TEST EVALUATION FORM--A

1. How do you feel about the length of time that the pictures for the problems stayed on the screen?

   Too long 26%
   About right 71%
   Too short 3%

2. In general, how do you feel about the length of time allowed for answering each problem?

   Too long 35%
   About right 63%
   Too short 2%

3. Which presented the test problems more clearly, the pictures in the film test or the written descriptions in the paper and pencil final?

   Written descriptions in paper and pencil test 9%
   Pictures in the film test 91%

4. How close did the film come to showing you parts, operations, and performances the way you saw them when you were going through the course?

   The film test showed things better than I saw them in the course 36%
   The film test showed things about the same as I saw them in the course 37%
   The film test showed things not as well as I saw them in the course 27%

5. Did the film test show you everything you needed in order to answer each problem? Yes 71% No 29%

   If you answered "no", check the suggestions you would make:
   More close-ups 49%
   Repetition of picture sequences 27%
   Better orientation shots 24%

6. How did the sound effects sound to you? Phoney 15% Real 85%

7. What did you think of the narrator's reading each problem and answer choices along with you?

It was a help 80%
Didn't make any difference 14%
Interfered with my own reading 6%

8. Do you think this film test taught you anything new about the material taught in the course?

Taught me a lot 29%
Taught me a little 62%
Taught me nothing 9%

9. Would you prefer a film test in which all problems relating to one subject or one piece of equipment were grouped together? Yes 63%
No 37%

If you answered "yes", check off the reasons why:
Such a test would be easier than the present test 38%
Would be more interesting 40%
Would be more realistic 22%
Other? _____

10. Which did you like better, the paper and pencil final or the film test?

Paper and pencil final 11%
Film Test 78%
Both the same 11%

11. Which did you find more interesting, the paper and pencil final or the film test?

Paper and pencil final 8%
Film test 84%
Both the same 8%

12. Did you stay awake throughout the entire film test? Yes 88%
No 12%

If not, why?
Was out late the night before 24%
Test was entirely too long 47%
Dark room made me sleepy 19%
Couldn't get interested 10%

13. Which test held your interest better?

The film test 80%
The paper and pencil final 8%
Both the same 12%

14. Which test gave you a better chance to show what you can actually do in the field?

Paper and pencil final 3%
Film test 82%
No difference 15%

15. The film test you took had over 200 problems. How long do you think this test ought to be to cover the course satisfactorily?

(Put in number) _____ Problems

16. What would you think about breaking this test up into grading period sections, using it as a grading period exam, and having a discussion on each problem after the students answer it?

Think it would help a great deal 79%
Think it would help somewhat 12%
Think it wouldn't make any difference 5%
Think it would be a waste of time 4%

17. Do you think you will have any use, in the near future, for the material you learned in the course?

Will have a great deal of use for it 53%
Will have some use for it 41%
Will have no use for it at all 6%

18. Make any other comments you wish:

APPENDIX K

CROSS REFERENCE FOR FILM PROBLEM NUMBERS AND PARALLEL AND COMPARABLE PAPER AND PENCIL ITEM NUMBERS

Paper and Pencil Test A and Film Test A

Item Number on Final School Examination	Problem Number on Film Test
22	76
24	25
36	161
37	173
38	54
43	113
46	47
55	41
62	104
67	105 (208)
68	189 (203)
82	72
84	185
94	32
115	140
120	141
126	31
127	120
131	50
136	27
138	13
139	18
144	132

Paper and Pencil Test B and Film Test B

Item No. on Paper and Pencil Test	Problem No. on Film Test	Item No. on Paper and Pencil Test	Problem No. on Film Test
1	89	28	7
2	94	29	137
3	92	30	12
4	14	31	86
5	122	32	93
6	124	33	147
7	23	34	167
8	126	35	181 (201)
9	127	36	159
10	24	37	196
11	95	38	155
12	99	39	183
13	58	40	182
14	59	41	92 (206)
15	28	42	171
16	3	43	200 (207)
17	36	44	198
18	5	45	156
19	101	46	184
20	116	47	181
21	68	48	177
22	118	49	153
23	37	50	149
24	6		
25	77		
26	52		
27	134		

APPENDIX L

GENERAL CONSIDERATIONS

1. Problem Selection

In selecting items to make two comparable forms of a film test, a number of criteria were considered. The statistical techniques used were the estimated biserial correlation of a problem with the criterion measure, and the difficulty of the problem.

In order to provide cross-validation of the items selected, the total group of subjects available for this project was divided into two parts. The biserial correlation and problem difficulty values were computed on Classes 69-74 inclusive.

The estimates of the biserial correlations were computed by finding the proportion of people in the upper and lower 37% of the distributions of the scores on the criterion variable. However, the people whose scores were in the upper and lower 4% on the criterion measure were given a weight of three, and the people whose scores were between the 5th and 16th percentiles inclusive or between the 84th and 95th percentiles inclusive were given a weight of two. The scores of the remainder of the people in the upper and lower 37% on the criterion measure were given a weight of one. This method differs from the more commonly used one of basing the biserial correlation on the answers of the people in the upper and lower 27% on the criterion measure, and was used because of the small number of cases (164) available for item selection. The biserial correlations computed by differential weighting of the extreme cases in the upper and lower 37% on the criterion measure have a smaller variance error than those computed by the more commonly used method, and therefore should be better estimates of the "true" correlation in the population from which this sample was drawn.[1]

The estimates of item difficulty were based on all of the people in Classes 69-74.

---

[1] The decision to use this method was based on work done by Dr. John C. Flanagan (3). The mimeographed charts used for this work were provided by Dr. Flanagan.

The following procedures were used for the final selection of the problems to be used in the film test. The two hundred problems available were reviewed jointly by three judges who rated the problems on a four point scale with respect to the importance of sound and motion for giving information necessary to answer the question. A problem was given a rating of four when in the opinion of the judges the motion and/or sound was crucial to answering the question. In general it may be said that a rating of four indicates that the information tested by that problem either could not be tested by the paper and pencil medium at all, or could not be tested without significantly changing or extending the premise and/or distractors. A rating of three indicates that sound and/or motion, while not crucial to answering the problem, made a significant contribution. The information tested by these problems could be tested with the paper and pencil medium but not without a complex premise and/or a number of still pictures, sketches or diagrams. A rating of two was given if the problem contained motion or sound which was superfluous. A rating of one was given if the problem was completely static. Problems rated two or one were thought by the judges to be capable of translation into the paper and pencil form, usually by a minor change in premise and the use of still pictures, sketches or diagrams. It should be noted that the difference between a rating of two and a rating of one is descriptive rather than qualitative.

After the problems had been rated in this way they were chosen for use in the final film test by five people working together, on the basis of the sound-motion rating, the problem correlations and difficulties, and in some cases, the number of subjects choosing each of the incorrect alternative choices. In this way, 110 of the 200 problems were chosen to be included in the film test, with a number of additional problems which might be included. In the problem selection, no problem was included that had an estimated biserial correlation less than .20. After these 110 problems, together with 10 of the best additional problems had been chosen, they were divided into ten areas according to the subject matter covered. Within these areas the items were divided into two forms, matched item by item as well as possible on the basis of the difficulty of the item and the correlation of the item with the criterion. These forms will be referred to as Film Tests I and II.

When the selected items had been divided into two forms in this way, the forms were examined to insure that one of them did not have two or more items that were alike, while the other had no items on this same material. Where it appeared that two problems were alike the problems were changed from form to form within a subject area so that the forms would still be as nearly equivalent in problem difficulty and biserial correlation as possible. In this process, two of the 120 items were dropped, so that the resulting comparable forms each had 59 items. Tests of this length were chosen because it was thought, on the basis of estimates made from the problem data for the first half of the group, that they would be long enough that the half tests would have a reasonably high reliability when given to a second sample. It seemed that tests much shorter than 60 items would not have high enough reliability to be useful for individual prediction where economically or personally important decisions were to be made.

It should be noted here that what has been done probably gives only a good approximation to two comparable tests. A commonly used definition of comparable tests is that they yield the same distribution of "true" scores, but it would seem necessary also that they cover almost exactly the same material, with items as much alike as possible. As the work for this contract was carried out eventually, it was not feasible to film enough problems to yield two comparable forms of a test that would correspond problem for problem in terms of material tested, problem difficulty, and correlation of the problem with the criterion measure.

APPENDIX L-2

CRITERIA FOR CONSTRUCTION OF FILM TEST PROBLEMS

1. It should be feasible to administer. Practical administration of the proposed test required that its format make it suitable to use with army projection facilities: a darkened room plus a projector and screen. Continuous response equipment, though developed to a practical stage, (1, 2) is not yet generally available on Army posts, and hence the design of film tests depending on availability of such equipment was not undertaken. Although there is good reason to believe that the optimum employment of the sound motion picture for testing purposes would include the use of continuous response equipment and the design of items and tests based on its characteristics for continuous or intermittent judgments by trainees as a film test proceeds, this approach was necessarily left unexplored for the reason stated above.

The assumed need to test relatively large groups of men at one time using regular available Army projection equipment ruled out the use of daylight rear projection which could have provided adequate viewing conditions and, at the same time, sufficient light in the room for writing responses. This daylight viewing equipment would permit an easy solution of the problem of conflicting physical requirements for viewing (relative darkness) and making responses (lighted room).

The rear projection type of screen being generally smaller than a regular screen limits the number of viewers and imposes some restriction on the type of material which can be adequately presented. Standard projection facilities, on the other hand, while providing excellent viewing conditions create the problem of providing adequate light for subjects to make responses. For standardization of the testing procedure, it was found to be practical for this light to be provided by the projector light reflected from the screen and controlled in duration by having clear film included in the test reels. This meant that the film test itself would provide and control the light for response periods of the test.

A pilot study indicated that under conditions of limited projection distance (30 feet) from a clean beaded screen, the light provided by a 1000 watt projection lamp, shining through clear film would provide adequate reflected light to allow subjects to record responses on a standard IBM sheet.

Later, in the particular testing situation at Aberdeen, it was found that the small amount of light used in the testing room (see p. 78) was sufficient to allow subjects to make responses to problems. This meant that the answer time provided by and controlled by the film itself was no longer necessary or effective as a control of when students would answer. However, since similar testing conditions could hardly be assured for all possible testing uses of the film, the clear film footage for light provision was left intact in each item.

In the planning stages of this project, the danger of "cuing" answers (cheating or "GIing" the test) was very seriously considered. It was proposed, for example, that if a multiple-choice type of problem form were used, and if the four alternatives were presented serially, i.e., one after the other in time, the response movements of members of the test population who knew the answers could "cue" other members as to correct answers. The "star pupil," simply by bending his head down to make a response while alternative number two was on the screen, could indicate purposely or inadvertently that he believed number two was the correct answer.

In the early stages of the project, therefore, the design of items using the serial presentation of alternatives was avoided. Later, a pilot study indicated that the danger of "cuing" was not as great as anticipated and some problems with the serial alternative form (i.e., answer alternatives presented pictorially in sequence) were included.

2. It should be efficient and objective to score. A basic requirement that the test be in such form as to provide for feasible and objective scoring as well as reliability considerations led to a decision to employ the multiple-choice form as the basic type for problems in the film test.

It should be pointed out, here, that lacking evidence on the relative merits of different film problem types and assuming the need to validate the proposed test against strict criteria, it was not considered advisable initially to attempt to create a test which, at the same time, embodied a complete range of as yet unevaluated film problem types and which also was presumably giving adequate coverage to a course of study. A decision was therefore made that the test problem pool (eventually 200 items including three sample problems) would consist of problems cast basically in a four alternative multiple-choice form. In addition, by way of demonstration, one longer "open-end" problem using a check-list answer sheet was developed as an adjunct to the main test; and a third major test problem was developed which called for trainee judgments based on the behavior and performance of four men doing a relatively complex job. A special answer form was designed for this test problem.

3. It should reduce demands on reading comprehension. Another important assumed requirement for the test was that it should reduce the demands on reading comprehension of the subjects. Since one of the features of the sound motion picture is its ability to present sound--voices and/or sounds--it was felt that a strong effort should be made to eliminate the long-standing objection to written tests, namely the demand on reading comprehension, by having a narrator read questions and alternative answers along with presenting the latter in printed form on the screen.

Furthermore, in order to control the period during which subjects might read alternatives it was decided that printed alternatives should appear on the screen where their appearance is controlled in time and order. It was planned that they should not appear on a printed sheet in the subjects' hands, because this condition would require divided attention between screen and test booklet. The subject might look at the booklet during a critical presentation on the screen.

It was planned, also, that the printed problem and alternative answers should appear before as well as following the picturization of the problem. In order to look at a motion picture sequence with the proper "problem set," with attention focused on the specific possibilities offered in the alternative answers rather than permit ad lib focus of attention on a multitude of details, it was believed that a statement of the main problem and alternatives prior to the motion picture sequence was necessary. These were to be read by the commentator while being shown in titles. It was assumed that this procedure, furthermore, would reduce the need for complete repetition of a problem and permit more efficient use of testing time. The alternatives were repeated after the presentation of the problem to reduce the need for remembering the choices.

4. It should not put trainees to sleep. Another assumed requirement for the test was that it should contain within itself the means for minimizing the "go-to-sleep" tendency typical of service men particularly in the film-viewing situation. By way of helping solve this problem a warning buzzer sound was planned to be used near the end of each test item. This warning buzzer was intended, also, to warn subjects that the answer period was coming to a close and a new problem would shortly begin.

Although the buzzer sounds were actually recorded on several hundred feet of sound track, their use was abandoned and those recorded were deleted because the cumulative effect of the buzzer sounds over a long period was judged to be annoying and distracting.

5. It should permit re-ordering of items without affecting difficulty levels. The pool of film problems as it was finally administered by the Instructional Film Research Program to the subjects at Aberdeen consisted of 200 test problems arranged in random order. The decision to randomize the items rather than present them as homogeneous groups dealing with areas of subject matter was made after careful consideration of the established merits of the latter form. It was believed that problem difficulty, which would be important in any final selection of problems for a test, would depend in some measure on the ordering of problems and the position of problems (beginning or end of test), and that a re-ordering in the final form might presume difficulty levels which no longer pertained. It was believed that the maximum consistency in problem difficulty would apply if problems were randomized in both the preliminary test and the final test made up of selected problems from the preliminary test.

The film test was also designed so that it could be rearranged simply by cutting and splicing a finished print. The sound track was timed so that the commentary always began at least two seconds after the picture. This makes it possible to cut the film between items.

APPENDIX L-3

THE FILMING OF THE TEST PROBLEMS

Introduction

Early inquiries indicated that the cost of producing test problems in motion picture form would be prohibitive if conventional professional 35mm. film production equipment, production methods and crews were used. It was therefore decided that the production should be done on the most economical basis possible, consistent with the achieving of acceptable film quality, and that camera equipment would be used which might be available to service units. It was also decided that production would use 16mm. film and equipment rather than 35 mm.

Film Production Personnel

Two film production teams A and B were formed. Each consisted of a camera man (with some experience in filming industrial subjects) and a director (with experience in film research) from The Pennsylvania State College. Camera team A had in addition a camera assistant.

Additional personnel were obtained from the Automotive Branch, Maintenance Training Division of The Ordnance School. To each team the following people were assigned: a Technical Advisor, generally a senior N.C.O. instructor from the School, whose function it was to see that technical accuracy and Army doctrine were adhered to in filming the problems. These technical advisors were changed as the teams moved from section to section in the School. In addition two army assistants (usually with the rank of PFC) were assigned to each team. The function of one of these assistants was to perform the necessary actions to be filmed in the test problems; the other assistant usually worked ahead of the camera team preparing tank equipment for filming. When necessary, additional personnel were obtained on the spot. Each camera team also had an "expeditor" from The Pennsylvania State College whose function was to obtain tanks for filming, arrange for assistants, and generally to pave the way for the production team.

The camera teams were made up of personnel from The Pennsylvania State College, and The Ordnance School as follows:

Team A	Team B
Camera Man (PSC)	Camera Man (PSC)
Assistant (PSC)	Director (PSC)
Director (PSC)	Expeditor (PSC)
Expeditor (PSC)	Technical Advisor (TOS)
Technical Advisor (TOS)	Assistant (TOS)
Assistant (TOS)	Assistant (TOS)
Assistant (TOS)	

Since not all of the test problems involved sound recording it was not necessary to have a sound engineer on location all of the time. The problems involving sound were filmed as a group during one two-week period. A sound engineer from The Pennsylvania State College assisted by a Director from one of the camera teams handled the recording.

Equipment

As was mentioned in the introduction to this section the equipment was kept as simple, light and portable as possible within the requirements of doing an adequate photographic job. The A and B camera teams had the following equipment:

Camera Team A

1 Eastman Cine Kodak Special 16mm. Camera with three lenses: 15mm., 25mm., and 63mm. The camera was spring driven.

1 Akerley Tripod and tripod leg holder

1 Norwood exposure meter

Lighting equipment as follows:

Two sets of "Color Tran" lights (one set consisted of three 150 watt PAR 38 lamps and one 300 watt flood together with a "Color Tran" voltage booster. The other set consisted of two 450 watt floods and "Color Tran" booster). (This team had, in addition, four 1500 watt floods, two 2000 watt spotlights, and two 750 watt spotlights, together with electrical equipment for balancing the load across several circuits. These lighting units were used only for lighting extensive areas, i.e., whole tanks which comprised about 2% of the shooting.)

In addition, this team had a Maurer 16mm. Motion Picture Camera available, but this was not used.

Camera Team B

1 Eastman Cine Kodak Special 16mm. Camera with three lenses: 15mm., 25mm., and 63mm.

1 Synchronous electric motor for camera

1 Quickset senior tripod

1 Eastman Kodak tripod dolly

1 Norwood exposure meter

2 "Color Tran" voltage booster transformers

8 Victor lamp stands each equipped with a 150 watt PAR 38 flood lamp

The equipment used by Team B is considered to be the minimum adequate for this type of basic film production. It was purchased especially for the job at a total cost of $1,275.67.

Sound Recording Equipment

For recording sound a Stancil Hoffman S5 magnetic recorder using 16mm. magnetic film was used. Since post-synchronization of picture and sound was satisfactory for most of the test problems, many of the sounds were recorded after the camera crew had filmed a particular problem. In the few problems where synchronized sound was required the synchronous drive motor was used on the Cine Kodak Special Camera, which then operated in synchronism with the Stancil Hoffman recorder. A directional microphone was used and sound proofing the camera proved to be unnecessary.

Use of the Equipment

Cameras

The Eastman Cine Kodak Special camera proved to be very satisfactory for this kind of production. For most purposes the spring motor was adequate since it will drive the camera for about 60 seconds. In a few instances where a shot required more than 60 seconds of continuous filming the electric motor drive was essential. The smallness and light weight of the Eastman Camera was a great advantage in working in cramped spaces and in moving around on tanks. Although Camera Team A had a Maurer Professional 16mm. camera available, it was not necessary nor feasible to use it. Thirty-five millimeter motion picture equipment would have greatly complicated production procedures since many shots had to be taken inside the tank turrets. The spring drive in the Cine Kodak Special camera was an advantage when working out in the field where power was not available.

Lighting

The eight PAR 38 lights operated through the Color Tran lighting transformers at a color temperature of 3200° Kelvin were adequate to light an area up to about 8' x 6' to an intensity of about 500 foot candles. This permitted a lens opening of f3.5 with Super X reversal film. (Super X 16mm reversal film was used throughout the work.) This kind of lighting equipment was adequate for most of the shots. It was lightweight, easily portable, and essential for working in confined spaces inside tank turrets. In such places the lamps were held in position with spring clamps. Another advantage was that these lights could be run directly from the regular power outlets in the building. Most of the shots covering large areas (e.g., entire tanks) were taken out-of-doors. In the few instances where large areas had to be filmed indoors, as in the power pack removal problem, some 2000 watt spot and flood lights were used.

Sound Equipment

The Stancil Hoffman magnetic sound recorder proved to be practical for location use, both in terms of recording quality, and portability. Sounds could be played back immediately, and since the magnetic film is driven by a synchronous motor and sprocket, exact synchronization with the camera could be obtained when necessary.

Procedure Used in Filming

The two camera teams worked in separate workshop buildings. The film test scripts for the problems to be filmed in a particular building were arranged according to the tank equipment to be used. One man worked ahead getting equipment and tools ready for filming while the rest of the team was filming a particular test problem. While the camera man was setting up his lights, the director and technical advisor would rehearse the operation to be filmed. Care was taken to ensure technical accuracy and adherence to Army doctrine in methods of work and dress.

It was found that each camera team could film an average of 10 separate film test problems in a day (from 8 a.m. to 5 p.m.). This included the time taken to set up and break down the camera equipment each day.

A record of each shot was made on the back of the film problem script sheets, and each shot was identified by photographing its number onto the film. Generally speaking, satisfactory results were obtained on the first "take," but where there was any uncertainty, or where the action was not quite right a second take was made. The ratio of film shot to film used in the finished production was about 1.5:1.

The greatest source of delay in shooting resulted from tank equipment which could not be made to operate in the desired way. In filming test problems liberal use was made of close ups. Care was taken to include all information necessary for the solution of the test problem. Judgment was exercised as to how much orientation, in the form of long shots, should be included. Ample running time was also allowed for each shot. Where the "distractors" (alternative answers) in a test problem were shown in the picture, they were given equal weight with the right response in terms of image size on the screen and running time.

Titles

It was decided in designing the test problems that the four alternative answer choices for each problem should be presented visually on the screen as well as spoken by the narrator. Also the problem number was to be included on a separate title. This required between 400 and 500 conventional title cards. The production of these cards would normally be a very costly item.[1]

---

[1] The estimated cost of title cards was $2200; the actual cost of linotype printing was $500.

In the interests of speed and economy, it was decided to have the titles set up in linotype by a local newspaper office, and to have a proof run off on good quality paper. Thus the titles were printed in lines up to 5" in length, on long strips of 6" wide paper. A small frame to line up each title was devised, and the filming of the titles proceeded very rapidly. The titles were underexposed slightly so that they came out as black letters on a grey background.

Special Test Problems

Two special types of test problems were attempted--one handled by each camera team. The first was a practical field evacuation problem which was filmed in some woods near the Aberdeen driving range. The second test problem covered the steps involved in the removal of the power pack from the M46 tank. This was filmed in a workshop building. Since a number of long shots showing the entire rear end of the tank were required, this problem taxed the lighting equipment to the limit.

Processing and Editing the Films

The film exposed each day was sent to a laboratory in Washington for reversal processing. A work print for editing was ordered with edge numbers stamped on each foot of the work print and original film to facilitate the editing process.

Editing of the work print was done at The Pennsylvania State College. The shots for each film problem were assembled in correct order, and the problems themselves were arranged on 1200' reels. Of the total of 17000' of film shot, 12000' was finally used. The 208 filmed problems were given an initial review by a panel at the Instructional Film Research Program. Suggestions were made for the rearranging and final editing of the test problems. A random order of presentation was worked out.

After this final editing the film problems were again reviewed, this time with the assistance of personnel from The Training Standards Division and The Automotive Branch at The Ordnance School and the AGO representative. A number of valuable suggestions were made, especially with reference to the wording of answer choices. Several of the 208 problems were subsequently eliminated for technical reasons. The number of problems finally tested was 200, of which three were sample problems.

Sound Recording and Printing

Commentary scripts were worked out and carefully timed to the pictures, and a narrator was selected from the faculty of The Pennsylvania State College.[2] The sound effects recorded earlier at Aberdeen had been transferred to sound tracks on film and these were mixed in with the narrator's voice in the final recording sessions. These combined recordings were made onto magnetic film using the Stancil Hoffman recorder. When satisfactory recordings were obtained they were transferred to film sound tracks. Meanwhile the original film was edited to match the final work print, and composite prints of picture and sound were made from a duplicate negative of the original film and from the sound track on film.

Film Costs

The actual cost of producing the 200 finished film test problems was approximately $20,000. This represents a cost of about $100 per problem. It should be borne in mind that this was the first time a test of this kind has been attempted. On the basis of the experience gained it is believed that the cost per problem on a future test could be reduced by about one-third.

Of course, the kind of content to be covered by the film test would greatly affect the cost, as some kinds of problems take much more time than others to set up and film. The cost of $20,000 for 12,000 feet of finished 16mm. sound film may be compared with an estimated cost of $100,000 (unofficially supplied by the Signal Corps) for production on 35mm. film using regular camera crews.

---

[2] The narrator was instructed to use a conversational delivery and to achieve the effects of sympathetic man-to-man, encouraging qualities of speech. Authoritative, pedantic and 'conventional' styles of narration were avoidéd.

APPENDIX L-4

ADMINISTRATION CONDITIONS FOR MOTION PICTURE TEST

1. Projection Room

A small theater seating 114 men was made available for administering the film test. The block of seats was arranged in ten slightly curved rows with more seats in the rear rows than in the front rows. All seats had folding desk arms for use in note taking. A sloping floor aided visi bility. The projector was enclosed in a booth at the rear of the room. This latter convenience was incidental to the administrative conditions and it is felt that projection with or without a booth would not critically affect the results.

2. Seating Arrangement

The groups tested varied in size from 20 to 50. The students were seated in alternate seats or in the case of small groups, in every fourth seat. Seating was arranged so that the students were sitting in a diagonal pattern. This maximized the distance a student had to look to read an answer sheet to his left front while the view of the answer sheet to his right front was obstructed by the back and head of the student sitting in front of him.

Seating Arrangement

O O X O O X O O X O O X O O

O X O O X O O X O O X O O X

X O O X O O X O O X O O X O

O = empty seat X = student

3. Screen Viewing Angle and Distance

The front two rows of seats were never used. This resulted in a viewing angle well within 30° from the center line. It is felt that all students had an excellent viewing angle. All students were seated between two picture-widths and five picture-widths from the screen. These viewing conditions are well within the limits established by AAF Aviation Psychology Program Research (4), as being necessary for adequate viewing.

4. Projector and Screen

The projector used was standard Signal Corps equipment (Projector Set AN/PFP-1). However, a Bell and Howell lens was substituted for the original lens. This resulted in a sharper and clearer picture around the edges. The throw was approximately six picture-widths (50 feet) so that all viewers were in front of the projector. A 1000 watt lamp was used.

The screen was the matte type and rather discolored. To this extent administrative conditions paralleled conditions that might be normally encountered. A cleaner screen or a beaded screen might have afforded more reflected light for the answer periods.

5. Room Illumination

Two shaded fluorescent lights at the extreme rear of the room were used for illumination during the test sessions. This light permitted easy use of IBM answer sheets but did not seriously reduce the perception of clear detail in the screen picture. This light also permitted easy proctoring of the room and tended to minimize the sleep-suggestive condition of a dark room. It was originally planned to use reflected light from the screen provided by a period of clear film for recording answers. Preliminary tests in a smaller room (30 foot throw) showed this procedure to be feasible. However, the larger room, the discolored screen, and the incidence of sleeping led to the utilization of a small amount of light in the room. Further work is necessary to determine the exact relation between room size (length of throw) and the use of projector-screen reflected light for recording answers. Black-out shades covered all windows.

6. Ventilation

Ventilation was probably the most unsatisfactory condition of the many administrative physical variables. However, it can be said that this testing took place under conditions probably normally found for projection (except in modern post theaters with forced draft ventilation). Several steps were taken to improve the situation.

a. Two rear doors were hooked in a slightly ajar position (6 in.) during the testing periods. On bright sunny days this introduced extra light which was compensated for by turning out the rear lights mentioned above.

b. Room heating was turned off during test sessions. Body heat warmed the room very quickly.

c. At every break period (between each reel) all students were asked to leave the room and go outdoors for fresh air. Four exit doors were opened for cross-ventilation. This resulted in maximum periods of 35 minutes that the room was closed, except for two rear doors slightly ajar, from one ventilating to another. With the low ceiling even this short period resulted in a "heavy" atmosphere particularly if payday was the day before testing!

7. Proctoring

A ratio of 1 proctor to 20 students was used. Proctors (from The Pennsylvania State College) would move about in the aisles and keep the test subjects under constant surveillance for signs of sleeping or cheating. No problem arose in the latter area.

APPENDIX L-5

ORDER OF PRESENTATION EFFECTS

An attempt was made to put as many of the film items in paper and pencil form as possible. These paper and pencil items were like the film items except that the stem of the item was worded to give the information that had been given in the picture or with sound, or both. The alternative answers were the same for both forms of presentation of the items. Because of the limitation on available testing time, only fifty of these items were presented to the men who had taken the film test. In this report these items will be referred to as Paper and Pencil Test B. These fifty items in film form will be called Film Test B.

In addition to these fifty items, twenty-three items which were thought to cover the same material as items on the film test were chosen from the final examination currently being used by The Ordnance School at Aberdeen. The judgments concerning comparability were made by two people working together. These twenty-three items will be referred to as Paper and Pencil Test A. These items in film form will be referred to as Film Test A.

Since the paper and pencil items in the two forms covered the same material as items on the film test, it was thought that the order of presentation of the paper and pencil and film items with respect to each other might be important. Therefore, before any comparison of the film items with paper and pencil items could be made, it was necessary to determine the effect of order of presentation. The students in Classes 69-74 inclusive answered Paper and Pencil Test B approximately 24 hours before they saw the comparable film items; those in Classes 75-77; 79-81 answered the comparable paper and pencil items approximately 24 hours after they saw the film items. The effect of order of presentation was tested by comparing the means and variances of the two groups on Paper and Pencil Test B and on Film Test B. The order of presentation of Paper and Pencil Test A could not be varied, since the twenty-three items were given as part of the final examination, which always preceded the film test. The final School examination was given the day before the film test was started.

The data for the comparison of the effect of order of presentation of the Paper and Pencil Test B with respect to Film Test B are presented in Tables 6 and 7.

TABLE 6

MEANS AND VARIANCES OF FILM TEST B
UNDER TWO ORDERS OF PRESENTATION

CLASSES 69-74 (Film Test After Paper and Pencil Test)		CLASSES 75-77; 79-81 (Film Test Before Paper and Pencil Test)	
$\bar{X}_1$	24.10	$\bar{X}_2$	22.35
S_1^2	52.0880	S_2^2	44.1218
$S^2_{\bar{x}_1}$	.3176	$S^2_{\bar{x}_2}$	.2902
N_1	164	N_2	152

C. R.$_{(\bar{x}_1 - \bar{x}_2)}$ = 2.24 p = .0125*

F = 1.18 p > .05

* A one tailed test of the means was used, because it had been predicted that the mean of the first group would be higher. The interest was not only in the magnitude of the difference, but also in the direction.

TABLE 7

MEANS AND VARIANCES OF PAPER AND PENCIL TEST B
UNDER TWO ORDERS OF PRESENTATION

CLASSES 69-74 (Paper and Pencil Test Before Film Test)		CLASSES 75-77; 79-81 (Paper and Pencil Test After Film Test)	
$\bar{X}_1$	20.75	$\bar{X}_2$	22.89
S_1^2	68.66	S_2^2	60.4800
$S^2_{\bar{x}_1}$	.4165	$S^2_{\bar{x}_2}$	.3979
N_1	164	N_2	152

$$C.R._{\bar{x}_2 - \bar{x}_1} = 2.38 \qquad p = .0087^*$$

$$F = 1.13 \qquad p > .05$$

* This is a one tailed test, since it had been predicted that the second group would have a higher mean than the first.

The items presented in the paper and pencil form were compared with the problems in film form by comparing the mean scores obtained on these items under the two media of presentation. The film test answer sheets were scored for the fifty comparable and twenty-three parallel items separately, and the comparisons were made separately, both for Paper and Pencil Tests A and B, and for Classes 69-74 and 75-77; 79-81. The method of presentation which yielded the higher mean score would be the easier method for answering, since the mean is just the sum of the difficulty indices for all the items.

The variances of the scores obtained under the two methods of presentation were also compared, because it was possible that the items could have the same difficulty values, but that the item intercorrelation might be different under the two methods of presentation. Because the data for the results of the two tests were obtained from the same people, these variances are not independent; therefore, it was not possible to make a precise test of the variances. They were tested using a critical ratio of the differences between the variances. The use of this test cannot be justified statistically because the differences in variance are not normally distributed. The square of the correlation between tests was used as the correlation between the variances. The desirability of making this test seemed to outweigh the weaknesses of the method.

Finally, the correlation of the scores obtained under the two methods of presentation with the average weekly practical rating were compared.

The data for the comparison of the paper and pencil items with the comparable or parallel items are presented in Tables 8, 9, and 10.

With both the Film Test and the Paper and Pencil Test, the mean of the test which was seen second was significantly higher than the mean of the test which was seen first. When Film Test B was seen after Paper and Pencil Test B, the mean score of the Film Test was significantly higher (C. R. = 2.24, p = .0125) than the mean score of the film test for the students who saw Film Test B before the Paper and Pencil Test B. In the same way, the mean score of Paper and Pencil Test B was significantly higher (C. R. = 2.38, p = .008) when it followed Film Test B than it was when it preceded Film Test B. Since there is data to indicate that the two groups of students used here could have been random samples from the same population, it appears that the order of presentation does make a difference, and that the comparison of the effect of the different media of presentation must be studied separately for the two orders of presentation.

TABLE 8

MEANS, VARIANCES, AND CORRELATIONS OF PAPER AND PENCIL TEST A AND FILM TEST A FOR TWO GROUPS OF CLASSES

CLASSES 69-74				CLASSES 75-77; 79-81			
Paper and Pencil Test A		Film Test A		Paper and Pencil Test A		Film Test A	
$\bar{X}_1$	10.50	$\bar{Y}_1$	12.09	$\bar{X}_2$	10.01	$\bar{Y}_2$	11.61
$S^2_{x_1}$	14.99	$S^2_{y_1}$	15.57	$S^2_{x_2}$	13.69	$S^2_{y_2}$	12.42
$S^2_{\bar{x}_1}$	.09	$S^2_{\bar{y}_1}$	.09	$S^2_{\bar{x}_2}$	.09	$S^2_{\bar{y}_2}$	.08
$S^2_{s^2_1}$	2.74	$S^2_{s^2_1}$	2.96	$S^2_{s^2_2}$	2.47	$S^2_{s^2_2}$	2.03
N_1	164	N_1	164	N_2	152	N_2	152

$r_{xy} = .61$

C. R. $_{\bar{x}_1 - \bar{y}_1} = 6.01$ $p < .001$

C. R. $_{s^2_{x_1} - s^2_{y_1}} = -.31$ $p = .76$

C. R. $_{\bar{x}_1 - \bar{x}_2} = 1.17$ $p = .24$

$r_{xy} = .62$

C.R. $_{\bar{x}_2 - \bar{y}_2} = 6.28$ $p < .001$

C.R. $_{s^2_{x_2} - s^2_{y_2}} = .76$ $p = .44$

$F = 1.09$ $p > .05$
163, 151

TABLE 9

MEANS, VARIANCES AND CORRELATIONS OF PAPER AND PENCIL TEST B AND FILM TEST B FOR TWO GROUPS OF CLASSES

CLASSES 69-74				CLASSES 75-77; 79-81			
Paper and Pencil Test B*		Film Test B		Paper and Pencil Test B*		Film Test B	
$\bar{X}_1$	20.75	$\bar{Y}_1$	24.10	$\bar{X}_2$	22.89	$\bar{Y}_2$	22.35
$S^2_{x_1}$	68.66	$S^2_{y_1}$	52.09	$S^2_{x_2}$	60.94	$S^2_{y_2}$	44.12
$S^2_{\bar{x}_1}$	.42	$S^2_{\bar{y}_1}$	.32	$S^2_{\bar{x}_2}$	.40	$S^2_{\bar{y}_2}$	.29
$S^2_{s^2_1}$	57.67	$S^2_{s^2}$	33.09	$S^2_{s^2}$	48.86	$S^2_{s^2}$	25.61
N_1	164	N_1	164	N_2	152	N_2	152
		r_{xy} = .78				r_{xy} = .86	
C. R. $\bar{x}_1 - \bar{y}_1$	= -8.17	$p \lesssim .001$		C. R. $\bar{x}_2 - \bar{y}_2$	= 1.69	p = .09	
C. R. $s^2_x - s^2_y$	= 2.70	p = .006		C. R. $s^2_x - s^2_y$	= 3.57	$p \ll .001$	

* After the Paper and Pencil Test B answer sheets had been scored, it was found that one of the items in this test had been made comparable to a problem which had been used as a sample problem in the Film Test, so that the Paper and Pencil Test B had one more item than the film test. The difficulty of this item for 132 people was .70. Consequently .70 was substracted from the mean of the Paper and Pencil Test B and .21 was subtracted from the variance of this test. This does not completely correct the variance so that it is equal to that which would have been found had this item not been scored, because of the intercorrelation of this item with all the other items. The correlation between the paper and pencil form and the film test form has not been corrected.

TABLE 10

CORRELATION OF PAPER AND PENCIL TESTS A AND B AND FILM TESTS A AND B WITH THE CRITERION MEASURE[1]

CLASSES 69-74		CLASSES 75-77; 79-81	
Paper and Pencil Test A	Film Test A	Paper and Pencil Test A	Film Test A
$r_{sc} = .71$ $z = .887$	$r_{tc} = .68$ $z = .829$	$r_{sc} = .62$ $z = .725$	$r_{tc} = .66$ $z = .793$
$N = 114$	$N = 114$	$N = 152$	$N = 152$
$C.R._{(z_{sc} - z_{tc})} = .66$	$p = .51$	$C.R._{(z_{sc} - z_{tc})} = .79$	$p = .43$
Paper and Pencil Test B	Film Test B	Paper and Pencil Test B	Film Test B
$r_{xc} = .68$ $z = .829$	$r_{yc} = .73$ $z = .929$	$r_{xc} = .71$ $z = .887$	$r_{yc} = .66$ $z = .793$
$N = 164$	$N = 164$	$N = 152$	$N = 152$
$C.R._{(z_{xc} - z_{yc})} = -1.47$	$p = .13$	$C.R._{(z_{xc} - z_{yc})} = 1.62$	$p = .11$

[1] To clarify the presentation of the data in this table, the following notation is used:

Paper and Pencil Test A is called (s)
Film Test A is called (t)
Paper and Pencil Test B is called (x)
and Film Test B is called (y)

The criterion measure is called (c). This notation applies only to this table.