DOCUMENT RESUME

ED 043 211 EM 008 195

AUTHOR Fields, Paul E.
TITLE Teaching Tests Revolutionize Effectiveness of Classroom Films.
PUB DATE Apr 70
NOTE 14p.; Paper presented at Annual Meeting of the Association for Educational Communications and Technology, formerly DAVI (Detroit, Michigan, April 27 - May 1, 1970)

EDRS PRICE EDRS Price MF-$0.25 HC-$0.80
DESCRIPTORS College Instruction, Culture Free Tests, *Instructional Films, Multiple Choice Tests, Programed Materials, *Student Testing, Testing, *Tests, Textbook Publications
IDENTIFIERS *Serial Multiple Discrimination Teaching Exam, Teaching Tests, TT

ABSTRACT

For instructional films to be effective teaching device, testing based on their subject material must be devised and used. Serial Multiple Discrimination Teaching Examinations (Teaching Tests or TT's) have been so used in an introductory psychology course attended by 5,000 students in the last five years. Since these tests have a low chance factor (the probability of getting a single question correct by guessing can be as low as 1/6,250), students do not guess. They study instead, and many of them have made perfect scores (40 consecutive errorless questions). Studies have shown that students taught by these examinations earn a higher proportion of correct answers and have a wider range of scores than those taught by other objective exams; they also scored above the mean on new exams over comparable material which were prepared by professors at other universities; and item analyses have consistently shown larger discrimination indices and higher internal consistency coefficients for TT's than for True False or Multiple Choice questions. (MF)

Teaching Tests Revolutionize Effectiveness of Classroom Films

Paul E. Fields

Professor of Psychology

University of Washington

Somehow, the bright future that was envisioned for A-V materials some 25 years ago has never materialized. One of the reasons for that failure may be laid directly to outmoded research studies which state that the approved method for using a film is to "introduce it, screen it, discuss it, and then show it again to answer all the questions." Instead of trying to find effective ways to help achieve the wider course objectives, these research studies attempted to justify the use of more film time, when, as evidenced by the proportion of film questions included on the final examination, their contribution was not important.

College textbook publishers have been forced to provide objective examinations to busy teachers as one of the chief inducements for using their books. Repeat adoptions are frequently lost, not because the text has gone out of date, but because the store of questions has been exhausted. If providing reliable and valid questions is essential to the achievement of specific educational objectives by college textbooks (to say nothing about increasing sales), why has no film ever been accompanied by a validated set of objective examination questions? It looks as if film producers have little faith in their product's ability to contribute to an approved educational objective.

But why has no further research been published? Without documenting it now, might I suggest that it is actually hazardous to a young teacher's career to do research on classroom teaching, and doubly dangerous when films are involved, because the threats come from both students and other faculty members. In addition, if his research findings are critical of the traditional views of the professional journals, few editors will publish them.

No matter how relevant, interesting, pedagogically sound, or technically perfect a film production may be, it is of no value unless it is used. Its use

must also be followed by significant changes in the desired direction in the measurable behavior of the learner. These changes cannot be determined without an additional expenditure of time and effort on the part of the teacher. The most important aspect of nearly any college course has always been the examination. Pressey ( ) demonstrated that regular examinations have positive teaching values, but Skinner ( ) has maintained that they are usually infrequent and aversive. Although I have successfully used different adaptations of the Pressey self-scoring teaching test techniques, the objectionable hardware was not eliminated until 40 years later when the principle of the latin square method was adapted for a paper and pencil test. Let me describe how we have been able to increase the value of films for more than 5,000 students during the last five years with the help of a new instructional device, the Serial Multiple Discrimination Teaching Examinations, or simply, Teaching Tests (TT).

Depending upon the seating capacity of the room, I now teach one introductory psychology class of 350 - 500 students each quarter; I cover the content of a larger, more complete text in 50 class sessions of one quarter than when the course extended over a full academic year; the level of achievement is now 50 to 100 per cent higher than it has ever been previously; and in addition, I now screen and examine over more than 50 reels of films, more than 100 slides of a 300 slide series, and more than 100 overhead transparencies. One half of all my exam questions are taken directly from A-V materials, with no loss of mastery of textbook content, and the modal score on most exams is a perfect score. Because of severe budgetary restrictions, I am assigned only one graduate Teaching Assistant, must use biweekly instead of weekly exams, and I spend 15 hours in student conferences each week.

Description of TT. A sample block of five TT questions and three special answer sheet formats are presented on page 1 of the handout. Each such block

of five TT questions has five numbered true-false stems with four lists of five true matching associations printed above them. On the special TT answer sheet, you are to slash the + or - symbol underneath the first question number to indicate whether you think the stem is true or false; then slash the one letter in each of the four columns of letters which corresponds to the closest true matching association in each of the four lists of five lettered matching alternatives. (The correct answers for questions 1 - 5 are slashed to illustrate the Verified Latin Square format, + E H L S +.) If you choose the correct matching alternatives, the symbol in column five opposite the row with no slashes will always verify your first tentative T-F answer to that question. Since each letter is only used once in a block of five questions, if the answers to any four questions are all correct, the answers to the fifth question must be the one letter and symbol in each column that has not been used previously.

Procedure. Students study TT questions of the VLS variety in their manual as a homework assignment before they see the film, and look for answers to the questions they had missed at home when it is screened in class. In order to eliminate rote memorizing, and emphasize the association of related ideas, class examinations are given over the same or new material with the ILS format (questions 6 - 10 show new orders for both questions and matching associations with the verification column eliminated) or the IRB format (in questions 11-15 the order of the matching associations has been randomized).

Teaching Tests over films have been administered under various combinations of before and after screening, and before and after having studied the questions in a verified latin square format. The amount and immediacy of feedback on the regular course exams was controlled and successively reduced by changing the order of the question parts, by using different VLS, ILS, and IRB blocks. In spite of the low chance success factor (the probability of getting a single question correct solely by guessing can be as low as 1/6,250) all five parts

must be correct to score a single point. Since the Teaching Test questions are prepared in blocks of five, if the student is not familiar with the TT technique and has no knowledge of the subject matter, the probability of answering any block of five questions correctly may be as low as 1 out of 9,536,813,164,062,500,000. If he is familiar with the technique but has no knowledge of the subject matter, the odds can still be as low as 1 out of 26,542,000 against him. With recognition that guessing simply cannot work in my classes, students do not try it. They study instead, and many of them have made perfect scores (40 consecutive errorless questions).

Results. Data from rows 1, 2, and 3 of Table 1 on page 2 reveal that new TT questions over lectures are harder than questions over the text or films; that new TT questions in row 4 are more difficult than old familiar questions in row 5; and that the same questions presented in randomized format are more difficult than in either VLS or ILS formats (rows 6 - 10). When the identical film questions were presented again in a new and more difficult format after seeing the films (rows 11 and 12), there was an absolute increase of 52 percentage points in achievement, with only an 8 percentage point drop in the discrimination index, and none in the internal consistency coefficient. This result is contrary to what has been found with other types of objective exam questions. In order to increase learning, all students should have ready pre-test access to all of the TT questions. The good students will master more questions at a faster rate than the slow students, resulting in high discrimination indices and high internal consistency coefficients, even though the average class achievement hits unprecedented levels.

Previous studies have shown that balanced experimental designs, where one student was given two types of questions or two formats on the same exam, always resulted in inflated t's because the students would spend most of their time on the question format that paid off with the highest scores. More conservative data were obtained when different groups of individuals were assigned to a single, but different, examination format on successive weeks. When data from many different TT questions of the same kind were summated across several quarters, an interesting new finding appeared (see Table 2). Scores for the top criterion group on both new and familiar textbook and film questions decrease with decreased feedback. Although scores for the bottom criterion group also decreased on new questions, they actually increased as less feedback was

provided on the old, familiar film questions. One possible explanation might be that the college students with poor verbal proficiency were able to manipulate the pictures more meaningfully than they could the verbal symbols. This finding is especially significant for the teaching of "disadvantaged" students. (The intercorrelations between the TT scores and the scores on the pre-college entrance test battery suggest that high "reading comprehension" scores are more important than so-called "intelligence", and that reading speed is not an important factor because it can be offset by high motivation and more study time.)

The data from Table 3 on the film Behavior Modification emphasizes the importance of feedback from the latin square format in answering complex questions. When the identical questions were presented again five weeks later (rows 2 and 4), but without the cues from the latin square arrangement or the verifying symbols, the top criterion group displayed the largest decline we had witnessed (95 to 86 %), but the bottom group increased from 36 to 41 per cent correct. A subsequent analysis revealed that it was impossible to discriminate logically between the matching associations for certain questions once feedback was removed. The only way they could be answered correctly on the retest was to remember which associations had been reinforced five weeks earlier with the latin squares. The discrimination index for these particularly difficult questions was higher than for the others (54. vs. 29.). Questions which had been included on an ILS exam (row 4) were answered with 8 percentage points higher accuracy than those which had not been included (row 5).

The new and unambiguous questions covering much simpler concepts taken from reviews of the Titicut Follies and the film (Table 4) were answered with excellent accuracy on their very first presentation (rows 1 and 4). With this relatively familiar and highly publicized subject matter, changing the examination format for new questions to reduce feedback was more detrimental to the

bottom group of students than to the top criterion group. In the last two rows (9 and 10) the top criterion group declined 1.8 percentage points while the bottom group lost 16.0 percentage points on the randomized format.

Earlier studies have shown that students taught with the TT's earn a higher proportion of correct answers, and have a wider range of scores than on other objective exams; they also scored above the mean on new exams over comparable material which were prepared by professors at other universities; and item analyses have consistently shown larger discrimination indices and higher internal consistency coefficients for TT than for TF or MC questions. Technically the TT's are not programed devices (although they can be), but the results from one early study suggested that they were as effective as certain programed texts. The availability of a programed film enabled us to extend our study to compare the relative effectiveness of the Fields Teaching Tests and the Markle and Tiemann Programed Notes in learning the content of the programed film Programing is a Process. (See page 5).

A conventional experimental design (Exhibit D) was selected, and the Programed Notes and the training TT's distributed to a selected half of the class. On the day after the film was screened a 30 item TT and a 120 item TF exam were administered. Unknown to the students, the TF exam contained 30 questions over each of four areas to help pinpoint experimental effects. The appropriate examinations were administered to four groups of 30 Psychology students and one group of 40 Engineering students.

Fifty-six per cent of the group who had studied the training TT's made perfect scores on the TT exam as opposed to seven per cent for the students who had studied the Program Notes. In spite of the very high average achievement level (85% correct) for those who had studied the training TT's, item analyses revealed significantly better average discrimination indices (.29 to .11) and internal consistency coefficients (.57 as opposed to .22) for the TT

than for the TF items (Table 5, rows 3 and 5).

Table 6 shows that among the Psychology groups, the total test scores of those students who had officially studied the TT's were superior to those who used the Program Notes at the .001 level of confidence. There was no significant difference between the total scores of the Psychology students and the Engineers who had studied only the program notes. When the part scores on the TF test were analyzed, the scores of the Psychology students who had studied the TT were significantly superior to those who had studied the Program Notes for all but the items which were taken directly from the film or notes and were not mentioned in the TT. Quite unexpectedly, the scores of the students who had studied the true matching associations in the training TT were significantly better for the 15 questions which had been changed from true to false. It was evident that these students had not rote memorized the TT's, but were responding to the larger ideas involved in the associations.

Conclusion. As long as films are designed by producers to appeal to as much of the kindergarden to adult spectrum as possible, and used by teachers to meet relatively low level entertainment objectives, they will not be utilized at anywhere near their true educational potential. But when films and Teaching Tests are used together, the quality of the learning closely approximates that obtained with a good textbook, and far exceeds the expectations of many producers. We now know how to use films in order to teach more effectively without the loss of precious classroom time. I hope that in the near future producers will provide us with films aimed at specific educational age levels, and filled with an abundance of up-to-the-minute data; that distributors will send complete scripts, and validated objective examination questions with each sale and rental print; that rental libraries will keep enough prints on hand to guarantee scheduling in the same week the textbook topic is assigned, if ordered six weeks in advance. The A-V industry and professionals must also help busy teachers measure the contribution of its own products to the learner.

Teaching Tests Revolutionize Effectiveness of Classroom Films

Paul E. Fields

University of Washington

Exhibit A. Illustrative Training Questions for Form E of Fields' Serial Multiple Discrimination Teaching Examinations.

A-Meijii Shrine, Tokyo	K-Tea and crumpets
B-Big Ben, London	L-Borsch and caviar
C-Eiffel Tower, Paris	M-Tempura and sake
D-Lincoln Memorial, Washington D.C.	N-Onion soup and crepes suzette
E-Kremlin, Moscow	O-Hot dogs and apple pie
F-Pennsylvania Avenue	P-Franc
G-Geishas	Q-Dollar
H-Red Square	R-Pound
I-Beatles	S-Ruble
J-Left Bank of Seine	T-Yen

1. The Union of Soviet Socialist Republics has the largest land area of any country.
2. France is the westernmost country in continental Europe.
3. Great Britain includes England, Wales, Scotland, and Ireland.
4. The capital of the United States of America is New York.
5. Japan has the world's most populous capital city.

Fields' Serial Multiple Discrimination Teaching Examination Answer Sheet. Form E

Print last name _____ First _____ Middle _____ Date _____ Class _____ Row _____ Seat No. _____

VLS	1-	A	F	K	P	/	2-	A	F	K	/	+	3-	A	F	X	P	+	4-	A	/	K	P	+	5-	X	F	K	P	+
		/	B	G	X	Q	-		+	D	G	L	Q	/		+	X	G	L	Q	-		+	B	G	L	X	-		/	B	X	L	Q	-
			C	X	M	R	-			/	E	H	M	R	-		/	C	H	M	X	-		/	C	H	M	R	/		-	C	H	X	R	-
		-	D	I	N	X	+		/	D	I	N	S	+		/	D	X	N	S	+		/	X	I	N	S	+			D	I	N	S	/
			X	J	O	T	-			E	/	O	T	-			E	J	O	T	/			E	J	X	T	-			E	J	O	X	-
	(3)						(2)							(5)					(1)						(4)										
ILS	6-	X	F	K	P		7-	A	F	K	/		8-	A	/	K	P		9-	A	F	K	P		10-	A	F	X	P						
		+	B	/	L	Q			+	X	G	L	Q			/	B	G	X	Q			/	B	G	L	X			+		B	G	L	Q	
		/	C	H	M	R			/	C	X	M	R				C	H	M	X			/	C	H	X	R			/	X	H	M	R	
			D	I	N	/				D	I	X	S			-	D	I	N	S			-	X	I	N	S				D	X	N	S	
			E	J	/	T				E	J	O	T				X	J	O	T				E	/	C	T				E	J	O	/	
	(4)						(1)							(2)					(5)						(3)										
IRB	11-	X	F	K	/		12-	A	/	K	P		13-	A	F	X	P		14-	A	F	K	P		15-	A	F	K	P						
		+	B	/	X	Q			/	X	G	L	Q			+	B	G	L	Q			/	B	G	L	X			+	B	G	L	Q	
		/	C	H	M	R				C	H	M	R			/	X	H	M	R				C	H	M	R			/	C	X	X	X	
		-	D	I	N	S			-	D	I	N	S			/	D	I	N	/			-	D	/	X	S			/	X	I	N	S	
			E	J	O	T				E	J	/	/				E	/	O	T				X	J	O	T				E	J	O	T	

Table 1. Item Analyses for SMDTE Questions over Text and Films.

Row	Materials.	Date	Form	No. of Ques.	Per cent passing H	L	Av.	DI	ICC
				New Questions					
1	Lect.		VLS	65	50	11	30	39	.52
2	Text		VLS	250	81	28	55	54	.62
3	Films		VLS	210	79	33	56	46	.56
				Questions from Teaching Test Manuals					
4	Text		VLS	370	96	42	69	54	.70
5	Films		ILS	240	96	53	75	43	.64
				Three Administrations of 30 "Narcotics" Questions					
6	Films	11/21/67	VLS	30	86	44	65	43	.51
7		1/9/68	VLS	30	58	18	38	40	.54
8		3/5/68	VLS	30	100	74	87	27	.51
9			ILS	30	100	68	84	32	.56
10			IRB	30	97	45	71	53	.65
				Pre- and Post-Screening Exams over 8 Films					
11	Films	4/30/68	VLS	40	50	16	33	34	.48
12		6/5/68	ILS	40	98	72	85	26	.50

Data under Per cent passing are for the two criterion groups; H is the average accuracy for the highest 27 per cent of the class, L is for the lowest 27 per cent. Av. = H + L/2.

DI = Discrimination Index, H - L.

ICC = Internal Consistency Coefficient, read from Table 3, Thorndike, Personnel Selection, 1949, page 348.

VLS = verified latin square; ILS = incomplete latin square; IRB = incomplete randomized block.

Exhibit B. Films Falling in Top 10 on All Three Criteria (Relevance, Teaching Effectiveness, and Interest) in Student Ratings.

June 1968 (38 films, 356 raters)

1. The Headbangers
2. Gateways to The Mind
3. Shades of Gray
4. Aggressive Child

December 1969 (38 films, 200 raters)

1. The Titicut Follies
2. The Headbangers
3. Behavior Modification
4. Detached Americans
5. Aggressive Child
6. Shades of Gray
7. Gateways to The Mind
8. Bright Boy, Bad Scholar

Table 2. Response to New and Old Questions over Textbook and Films under Three Teaching Test Formats.

	No. Ques.	Per cent Pass H	L	Av.	DI	ICC	No. Ques.	Per cent Pass H	L	Av.	DI	ICC
		Textbook						Films				
						VLS Format						
New Q.	100*	94.2	56.4	75.3	37.8	.54	110	91.9	50.3	71.1	41.6	.57
Old Q.	200	99.5	61.6	80.5	37.8	.60	110	98.7	41.4	70.1	57.3	.73
						ILS Format						
New Q.	175	81.6	27.6	54.6	54.0	.62	70	84.7	47.0	65.8	37.7	.53
Old Q.	160	98.9	46.9	72.9	52.0	.71	150	96.5	45.7	71.1	50.9	.68
						IRB Format						
New Q.	110	66.8	18.8	42.8	48.1	.59	22	75.7	23.8	49.7	41.9	.61
Old Q.	155	91.6	42.7	67.1	48.9	.65	130	93.7	53.8	73.8	39.9	.58

*Each question answered by minimum of 157 students.

Table 3. Item Analyses for Identical Film Questions under Three Exam Formats with Different Amounts of Feedback.

Behavior Modification (Complex Concepts)

Date	Format	No. Ques	Per cent Pass H	L	Av.	DI	ICC	Conditions
10/31/69	VLS	20	35.7	8.1	22.0	27.6	.49	New questions before seeing films; did not count.
11/7/69	ILS	10	95.0	36.4	65.7	53.6	.70	After VLS and seeing film; half on each W or Y form.
12/17/69	IRB	20	82.6	36.4	57.9	46.5	.54	Five weeks after VLS and seeing film.
			Effect of Extra ILS Exam on IRB Scores					
VLS ILS IRB 10/31 11/7 12/17		10	86.4	40.8	63.6	45.6	.53	Half on each W or Y form.
VLS IRB 10/31 12/17		10	79.4	32.0	55.7	47.4	.56	Half on each W or Y form.

Table 4. Item Analyses for Identical Film Questions under Three Exam Formats with Different Amounts of Feedback.

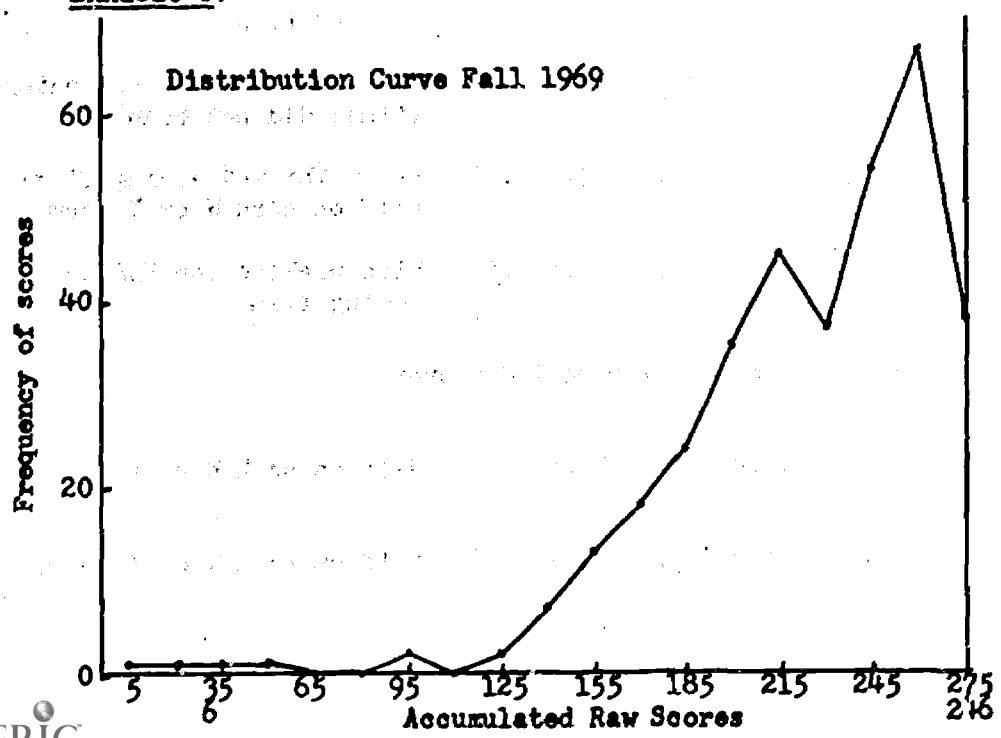
Reviews of Titicut Follies (Simple Concepts)

Date	Format	No. Ques	Per cent Pass H	L	Av.	DI	ICC	Conditions
11/10/69	VLS	10	95.4	45.6	70.5	49.8	.66	New questions before seeing films; did not count; no time limit.
12/17/69	ILS	10	91.4	69.4	80.4	22.0	.41	Five weeks after VLS; time limit. Half on W and Y form.
12/17/69	IRB	10	89.0	51.0	72.5	33.0	.47	

Titicut Follies Film (Simple Concepts)

11/21/69	VLS	20	96.7	75.6	86.2	21.1	.43	New questions after seeing film; counted on grade.
12/17/69	ILS	20	97.3	74.4	85.9	22.9	.48	Half on W and Y form. Counted.
12/17/69	IRB	20	95.5	65.7	80.6	29.8	.50	Half on W and Y form. Counted.
VLS 11/21	ILS 12/17	10	99.0	77.2	88.1	21.8	.50	Half on W and Y form. Counted.
VLS 11/21	IRB 12/17	10	97.6	75.6	86.6	22.0	.46	Half on W and Y form. Counted.
	ILS 12/17	10	95.6	71.8	83.7	23.3	.45	New questions; half on W and Y forms.
	IRB 12/17	10	93.8	55.8	74.8	38.0	.54	New Questions; half on W and Y forms.

Exhibit C.



Fall Quarter Grades

234	A	149
186	B	127
138	C	51
89	D	14
88-	E	7
	E/W	5
	Inc	6
	P/W	6
	Total	365

Exhibit D. Learning about "Programing is a Process" from "Programed Notes" vs. "Teaching Tests".

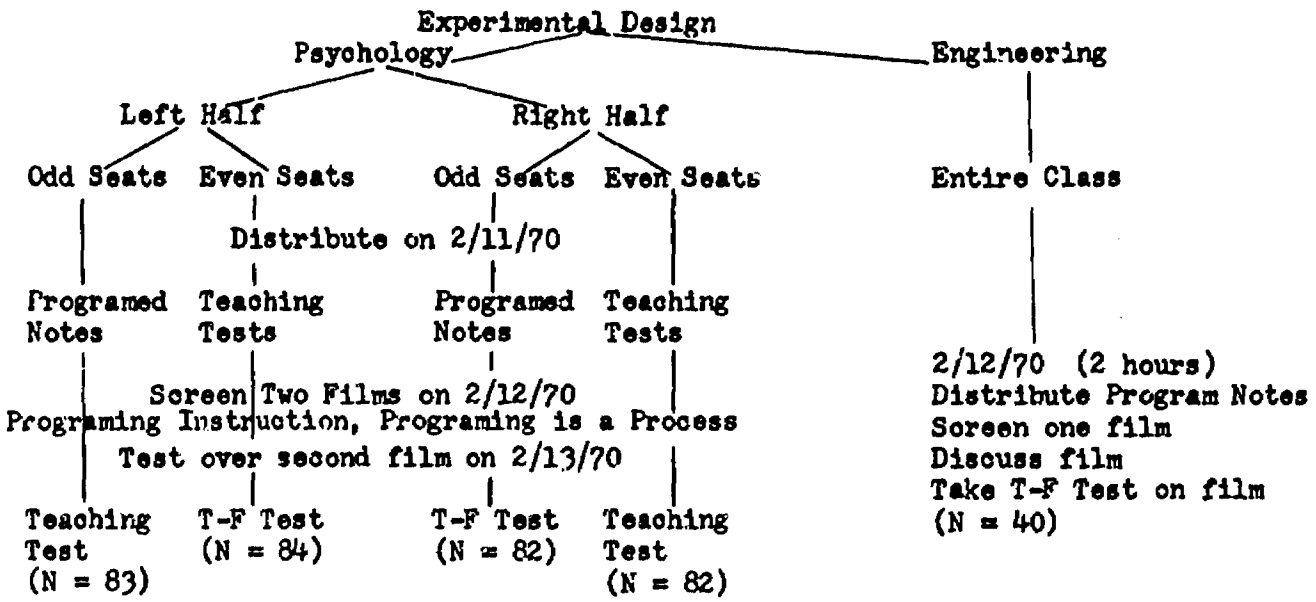


Table 5. Average Item Analyses of Total Test Scores for 120 T-F and 30 TT Items.

	Per cent passing				
True-False Test (No. right)	H	L	Av.	DI	ICC
Engineers, Programed Notes	85.50	73.08	79.29	12.42	.23
Psychology, Programed Notes	87.08	71.17	79.13	15.91	.27
Psychology, Teaching Tests	90.30	79.46	84.98	11.04	.22
Teaching Test (No. right)					
Psychology, Programed Notes	94.50	42.33	68.42	52.17	.67
Psychology, Teaching Test	100.00	71.00	85.00	29.00	.57

Table 6. Significance of the Difference between Means.

Total Scores	Psychology	Engineers
Teaching Test (Programed Notes vs. Teaching Tests)	TT 6.66***	
True-False Test (Programed Notes vs. Teaching Tests)	TT 6.86***	
(Programed Notes, Psychology vs. Engineers)		.65
Part Scores on T-F Test		
TT - Programed Notes Vs. Teaching Test	TT 3.55***	
M -	TT 9.49***	
MT -	TT 2.60*	
MF -	TT 4.80***	
Film -	.73	
New -	TT 2.20*	
TT - Programed Notes, Psychology vs. Engineers		.79
M -		1.14
MT -		.02
MF -		1.24
Film -		E 3.35**
New -		.93

Serial Multiple Discrimination Teaching Examination over Teaching Tests.

(IRB Format. Use key for items 11 - 15 at bottom of page 1)

A-Each question has a true-false sentence for its "stem"; the subject of the sentence should be noted.
B-Many students had developed "entertainment sets" which interfered with the learning of the basic factual content of the film.
C-With TF questions, the probability of a correct answer by chance is .5 or $\frac{1}{2}$.
D-An item analysis requires that the per cent of students passing each question be computed for the top and bottom criterion groups.
E-The matching associations for one question have been ordered so as to fall at one intersection of each row and column (latin square).

F-Students reported that films covered so many things and "went so fast" that it was impossible to take satisfactory notes.
G-One phrase or sentence in each of four lists of matching associations is closely related to the key idea in each true-false statement.
H-The difficulty of the question, its discrimination index, and the internal consistency coefficient are then determined.
I-If the first four matching associations for each question are correct, the + or - symbol in the fifth column opposite the unused row is the correct TF answer.
J-One formula frequently used to determine how much is known for sure is, "score = number right minus the number wrong."

K-With four or five-alternative Multiple Choice questions, the chance factor is reduced to .25 or .20 respectively.
L-Each of the sentences in the lists of matching associations is always true.
M-Even though achievement goes up significantly with the use of Teaching Tests, they still discriminate between good and poor students better than other tests.
N-If any four questions in the same block of five have been answered correctly, the unused letters and symbol must be the answers to the fifth question.
O-A critical and rational approach was sometimes difficult to develop because of over-dramatization and identification with particular actors.

P-Even though a true-false statement may be false, it must always be associated with four true associations in order to obtain a correct answer.
Q-If any letter or symbol has been used twice in the same block, another was omitted. A study of the duplicated and omitted parts should pin-point the error.
R-If larger proportions of the classes trained with Teaching Tests than by traditional methods consistently achieve the criterion, the TT must be an effective instructional device.
S-The probability of marking all five parts of the easiest TT question correctly without any knowledge is only 1 out of 3,125.
T-By studying and taking the Teaching Tests over films at home, students knew what to look for when the films were screened in class.

1. Description. A single Teaching Test question is composed of five separate true-false statements.
2. Purpose. One of the objectives for developing Teaching Tests over films was to make them teach effectively without using a disproportionate share of the total class time.
3. Guessing. The "guessing factor" has been minimized or well controlled in the preparation of traditional objective examinations.
4. Verification. Good students use "feedback" from the TT special answer sheet as they work through a VLS homework training assignment.
5. Statistical techniques. Once they have been released to students, TT questions over films become so easy that they are useless for assigning grades.