REPORT - SDC 269-7-7

RELATIONSHIP OF LENGTH AND FACT FREQUENCY TO EFFECTIVENESS OF INSTRUCTIONAL MOTION PICTURES

(Rapid Mass Learning)

The Pennsylvania State College
Instructional Film Research Program
November 1949

Project Designation NR-7
Contract N6onr-269, T.
SDC Human Engineering Project

Report R

TECHNICAL REPORT - SDC 269-7-7

RELATIONSHIP OF LENGTH AND FACT FREQUENCY TO EFFECTIVENESS OF INSTRUCTIONAL MOTION PICTURES

(Rapid Mass Learning)

The Pennsylvania State College
Instructional Film Research Program
November 1949

Project Designation NR-781-005
Contract N6onr-269, T.O. VII
SDC Human Engineering Project 20-E-4

Report Prepared by
W. S. Vincent, P. Ash, L. P. Greenhill

- -

FOR THE PENNSYLVANIA STATE COLLEGE:

Dean M. R. Trabue
Responsible Administrator

C. R. Carpenter
Program Director

- -

FOR THE SPECIAL DEVICES CENTER

Reviewed for Human Engineering Branch:

J. Gaberman, Project Engineer

C. P. Seitz, Head, Code 912

Submitted:

Fred W Dittler Cdr USN
for C. S. Rhoads, Technical Director

Approved:

C W Gordon, Cdr, USN
for J. R. Ruhsenberger, Captain, USN
Commanding Officer and Director

SUMMARY

The Problem

The purpose of this research is to determine the effect on learning of varying (1) the total amount of factual information presented in a film of a given length, and (2) the length of time allotted to conveying a fixed amount of information. The question is: Does increasing the concentration of facts in a film result in a proportionate increase in learning?

Procedure

Four experimental film versions dealing with the causes and manifestations of the weather were made up from a series of Navy training films on aerology. The Long Heavy version ran 29 minutes and contained 224 facts; the Long Light version also ran 29 minutes but contained 112 facts. The Short Heavy version ran 14 minutes and contained 112 facts; the Short Light version ran 14 minutes but contained 56 facts. The total number of words in each pair of equal length kept constant by the use of repetitions, prefatory statements, and other "filler" material which did not add new facts.

The four experimental versions were shown to four groups in each of three different populations, High School students (12th grade), Air Force basic trainees, and College students. In each population a fifth control group did not see the film. All groups took the same 136 item multiple-choice question information test. The High School and Air Force groups took the test again after delays of four weeks and seven weeks respectively.

Results

Significant learning occurred. Every group which saw experimental film earned a substantially higher score than the control group which did not see a film. The "best" version in an all-around sense on the basis of total score differed from population to population. For the High School sample the Short Heavy version seemed to be the most effective, for the Air Force and College samples the Long Light version seemed to be most effective. At the end of the delayed recall period all differences among the versions were much smaller than they had been on the immediate retention test, and most of them were not significant.

Conclusion

It seems clear from the data that packing more and more information into a film yields only very slight increments in total measured learning. In no case did the Long Heavy film group learn anything approaching twice as much as the Short Heavy or Long Light groups, nor did the latter learn twice as much as the Short Light group.

Analysis of the test performance suggested that the films were rather difficult for the populations used in spite of the fact that the reading level of the scripts was at the 7th or 8th grade.

RELATIONSHIP OF LENGTH AND FACT FREQUENCY TO EFFECTIVENESS OF INSTRUCTIONAL MOTION PICTURES

W. S. Vincent*, P. Ash and L. P. Greenhill

STATEMENT OF THE PROBLEM

The purpose of this research is to determine the effect on learning of varying (1) the total amount of information presented in a film of a given length, and (2) the length of time allotted to conveying a fixed amount of information. The experimental question posed is: Does increasing the fact density of a film result in a proportionate increase in the learning accomplished?

EXPERIMENTAL DESIGN AND PROCEDURES

The Films

Using as source material a series of films on aerology[1], visual material was selected for inclusion in four versions of an introductory film on the weather.

"The Weather" covered, in more or less detail depending upon the version, the basic facts with respect to the formation and characteristics of frontal weather, and the effect of weather conditions on flying.

A careful content analysis of a tentatively selected body of material was made to permit controlling within narrow limits the content of four versions of the film.

The unit of content employed was the individual

---

* Dr. W. S. Vincent was the initial project leader. Extensive work on this research was done by the Film Research Staff

1 The source films were in color and in animation. They had been produced by the Walt Disney studio for use by the Department of the Navy. The films included:

Aerology - Fog (MN-119B),
Aerology - Air Masses and Fronts (MN-119D),
Aerology - The Cold Front (MN-119E),
Aerology - The Warm Front (MN-119F),
Aerology - The Occluded Front (MN-119b),

fact, and a fact was defined as any item about which a question could be asked. This was designated as the "questioning to exhaustion" technique of testing. To identify all the facts in each script, a committee of eight or nine IFRP Staff Members read preliminary drafts, and wrote questions for every item mentioned. The scripts were modified so that the number of facts in each version could be carefully specified. The scripts for the two films of each length included facts in the ratio of 1:2. However, the total number of words in each pair of equal length was kept constant by the use of repetitions, prefatory statements, and other filler material which did not add new facts. Illustrations and examples were considered as repetitions. This material was included in both the visuals and the commentary. One version, designated as Long Heavy, included all the facts used. This version ran 30 minutes. A second version, the Long Light, included half the facts used in the Long Heavy but also ran 30 minutes. A third version, the Short Heavy, included all the facts that were in the Long Light version, but ran 15 minutes. Finally, a fourth version, the Short Light, also running 15 minutes, included only half the facts found in either the Short Heavy or the Long Light version.

To ensure further that the commentaries of the four versions were of equal verbal difficulty, and that the level of verbal difficulty was appropriate for twelfth grade high school students or military trainees of equivalent education, an analysis of the reading level of the four scripts was made. The Dale-Chall formula[2] was used for this purpose, and minor changes were made in the commentaries to obtain equality of reading difficulty. The formula is based on two counts: average sentence length, and percentage of unfamiliar words.

Table I summarizes the characteristics of the scripts for the four versions. The four versions were in color animation and in sound.

The Tests

An objective-type test employing four-choice questions was constructed. The questions used were those formulated to identify the facts in the films. Since a test of 224 items was considered too long, a sample of 136 of the questions was selected. The distribution of question coverage for the versions is also given in Table I. The same test was used for all groups.

---

2 Dale, E., Chall, J. S. Formula for predicting readability. Ed. Res. Bul., 1948, 27, 11-20, 37-54

TABLE I

CHARACTERISTICS OF THE EXPERIMENTAL VERSIONS OF "THE WEATHER"

	Version			
	Long Heavy	Long Light	Short Heavy	Short Light
Running Time	28.8 min.	29.1 min.	14.3 min.	14.3 min.
Total Number of Facts	224	112*	112*	56
Facts per minute	7.77	3.85	7.85	3.91
Total Number of Syllables	3599	3596	1745	1760
Syllables per minute	124.9	123.7	122.3	122.8
Verbal Difficulty of Script				
Dale-Chall Score	6.8071	6.8911	6.9652	6.7320
Grade Level	7-8	7-8	7-8	7-8
Number of Items in Test	136	89**	89**	45

* Same facts

** Same items

Experimental Procedures

The general procedure followed involved showing each of the four versions to one of four comparable groups, and testing these four groups and a comparable fifth Control Group which was not shown a film. The mean scores were compared for (1) the entire test, (2) for the items common to all but the Short Light version, (3) for items common to all versions, and (4) for items only in the Long Heavy version.

Three replications were conducted. The replications are summarized in Table 2.

The first replication employed twelfth grade students in the Lewistown, Williamsport, and Sunbury (Pennsylvania) high schools. In each school five groups were used. The high school students were tested for immediate recall and for four-weeks delayed recall.

The second replication employed ten flights of recruits (basic trainees) in the Air Force (Lackland Air Force Base, San Antonio, Texas). Each film version was shown to two flights, who were tested immediately and again seven weeks later.

The third replication employed students in five sections of a course in elementary meteorology at The Pennsylvania State College. These subjects were tested once only, one week after the film showings.

TABLE 2

SUMMARY OF THE EXPERIMENTAL REPLICATIONS

	Replication		
	High School	Air Force	College
Number of Subjects	434	513	324
Character	Male and female, 12th grade students, five groups in each of three high schools.	All male, ten intact companies of Air Force basic trainees (two companies to each treatment)	Male and female, in five intact sections of an introductory course in Meteorology
Date of Study (1949)	April-May	June-August	September-October
Treatment:			
Films	Yes (except for control group	Yes (except for control group)	Yes (except for control group)
Retention Test	Immediately after film	Immediately after film	One week after film
Delayed Recall Test	Four weeks after film*	Seven weeks after film	None

* Delayed recall test not readministered to Control Group.

The Populations

The distributions on available criteria for the three populations used in the study are given in Table 3.

For the high school students, only sex distribution data was obtained. In each high school, students were taken from their classes and distributed among the five treatment groups so as to ensure more or less comparability with respect to sex, course in which they were enrolled, and similar factors. The groups, as finally constituted did not differ significantly with respect to sex distribution (Chi-square not significant at the 30 per cent level).

For the Air Force basic trainees, only educational level was obtained. The methods of company formation employed in the Air Forces are such that one may be reasonably confident that each intact company (flight) is a random sample of the whole Air Force basic recruit population. Therefore, intact flights were used, without further randomization. The treatment groups (two flights each) did not depart significantly from homogeneity with respect to educational level (Chi-square not significant at the 10 per cent level).

For the College students, sex, semester, and curriculum data were available. Intact classes had to be used for the four film groups. The Control Group comprised a fifth class plus those students in the other four classes who were absent when the films were shown. It may be noted that the treatment groups, as here defined, departed significantly from homogeneity with respect to sex distribution and curriculum distribution (Chi-square significant at the 5 per cent level or better in both cases).

TABLE 3

DISTRIBUTION OF SUBJECTS BY SEX, SEMESTER, CURRICULUM AND/OR EDUCATION FOR THE COLLEGE, HIGH SCHOOL, AND AIR FORCES POPULATIONS

COLLEGE POPULATION

Group	Sex[1]		Semester[2]			Curriculum[3]			Total
	M	F	1-4	5-6	7-8	Sci.	Ag.	LA	
Long Heavy	27	6	5	11	17	13	8	12	33
Long Light	65	13	6	18	54	32	19	27	78
Short Heavy	38	8	9	12	25	11	12	23	46
Short Light	80	2	14	18	50	16	20	46	82
Control	73	12	17	21	47	15	22	48	85
TOTAL	283	41	51	80	193	87	81	156	324

HIGH SCHOOL POPULATION

Group	Sex[4]		
	M	F	
Long Heavy	40	40	80
Long Light	43	48	91
Short Heavy	34	49	83
Short Light	32	49	81
Control	50	49	99
TOTAL	199	235	434

AIR FORCE POPULATION

Group	Education[5]			
	Grade School	High School	College	
Long Heavy	15	92	0	107
Long Light	12	94	1	107
Short Heavy	9	93	5	107
Short Light	19	75	2	96
Control	13	71	2	86
TOTAL	68	425	10	503

1. Chi-square = 10.89, $.05 > P > .02$
2. Chi-square = 8.16, $P > .30$
3. Chi-square = 18.78, $.05 > P > .02$
4. Chi-square = 3.60, $P > .30$
5. Chi-square = 4.80, $P > .10$

RESULTS

The means for the film test scores for the groups seeing the four versions and for the Control Group are reported in Tables 4 (High school students), 5(Air Force basic trainees), and 6 (College meteorology students). For each of the five groups, means and related statistics are given for the following scores:

V_1 Score - based on 47 items covered by the Long Heavy version only. Mean scores for the groups seeing the other versions, where these mean scores were higher than the Control Group means, may be attributed to inferences.

V_3 Score - based on the 44 items common to the Long Heavy, Long Light, and Short Heavy versions. These items were not covered in the Short Light version.

V_4 Score - based on the 45 items common to all four versions. This score represents a measure of direct learning on all the tested material in the Short Light version.

T_2 Score - this is the sum of the V_3 and V_4 scores. It is based on the 89 items covered in the Long Light and Short Heavy versions.

Total Score - based on all the 136 items. This score covers all the information included in the Long Heavy version.

The tables of differences among the versions will not be included in this report, but they will be summarized briefly.

The following findings may be stated:

1. Significant forgetting took place. For both the high school sample (4-week interval) and the Air Force sample (7-week interval) the delayed recall test mean scores were about one standard deviation lower than the immediate recall means, and this difference was, in almost all cases, significant at the 0.1 percent level of confidence. The anomalous finding (Table 5) that significant "forgetting" took place in the Control Group in the Air Forces (this group did not see a film and, theoretically, learned nothing to forget) may be explained on the basis of very poor motivation on the recall test. This second administration of the long test presented the Control Group with an extremely frustrating task for the second time. On the second occasion, the group largely "gave up" and answered randomly.

2. Significant learning took place. For both the immediate recall test and the delayed recall test, for all three populations, almost every film group mean score is substantially (more than one standard deviation) and significant-

ly (at the 0.1 percent level of confidence) greater than the comparable Control Group score. The only exceptions were as follows: the V_1 and V_3 delayed recall means for the Short Light group in the high school sample were not significantly different from the high school V_1 and V_3 means for the Control Group; and the V_1 and V_3 means for the Short Light group in the College population were not significantly different from the College Control Group means for these scores. Since the V_1 and V_3 scores pertain to information not shown to the Short Light group, this finding is not surprising.

3. Some inferential learning took place. This is the converse of the finding reported above. Although not actually shown the items entering into the V_1 Score, the members of the Long Light and Short Heavy groups in all populations earned higher scores than the comparable Control Groups did, for both immediate and delayed recall tests. Furthermore, with the exceptions noted above, the Short Light group inferred significantly more V_1 and V_3 items than did the comparable Control Groups.

4. With regard to the inter-version comparisons, the following comments seem justifiable:

a. The "best" version, in an all-around sense, on the basis of the total score, differed from sample to sample. For the High School sample, the Short Heavy version seemed the most effective. For the Air Force and the College samples, the Long Light version seemed to be most effective.

b. In the Air Force and High School samples, the Long Heavy group scored significantly higher than any other on the V_1 score for the immediate recall test. At the end of the delayed recall interval, however, this difference approached zero, and was not significant in a statistical sense. In the College sample the Long Heavy group had a higher V_1 score than any other group at the end of one week, but only the difference from the Short Light group was statistically significant.

The V_1 score covered items included explicitly only in the Long Heavy version.

c. In general, the Short Light Group scored higher on the V_4 score (items common to all versions, and the only items in the Short Light version) than any other group. These differences were not large, however, and only a few were significant at the 5 percent level or better.

d. At the end of the delayed recall period, all differences among the versions were much smaller than they had been on the immediate retention test, and most of them were not significant.

TABLE 4

MEANS, STANDARD DEVIATIONS, AND STANDARD ERRORS OF MEANS FOR IMMEDIATE RECALL AND DELAYED RECALL SCORES, FOR HIGH SCHOOL POPULATION

Group	No. of cases	Immediate Recall Mean	S.D.	SE_m	Delayed Recall Mean	S.D.	SE_m	Mean Diff.	r
Total Score									
L H	80	62.2	16.2	1.82	50.1	15.4	1.73	12.1***	.77
L L	91	63.0	17.3	1.82	52.9	15.8	1.67	10.1***	.81
S H	83	64.1	15.8	1.74	54.7	15.4	1.70	9.4***	.90
S L	81	58.1	15.0	1.68	47.2	14.2	1.59	10.9***	.80
C	99	42.1	7.9	.80					
T_2 Score									
L H	80	41.9	11.5	1.29	33.8	11.0	1.24	8.1***	.61
L L	91	45.6	13.4	1.41	36.9	11.9	1.26	8.7***	.76
S H	83	46.6	12.6	1.39	38.1	11.5	1.27	8.5***	.88
S L	81	41.2	11.3	1.27	32.4	10.5	1.17	8.8***	.79
C	99	28.1	5.7	.57					
V_1 Score									
L H	80	20.4	5.4	.61	16.3	5.6	.63	4.0***	.67
L L	91	17.5	5.0	.52	16.0	4.8	.50	1.4***	.71
S H	83	17.4	4.2	.46	16.6	4.5	.49	.8*	.56
S L	81	16.9	4.6	.51	14.8	4.6	.51	2.1***	.53
C	99	14.0	3.4	.35					
V_3 Score									
L H	80	20.5	6.4	.72	16.7	5.7	.64	3.8***	.71
L L	91	22.3	6.5	.69	17.9	5.3	.56	4.5***	.71
S H	83	23.5	6.8	.75	18.4	6.3	.70	5.1***	.81
S L	81	16.7	5.3	.59	14.4	5.0	.56	2.3***	.67
C	99	13.9	3.4	.34					
V_4 Score									
L H	80	21.4	5.9	.67	17.1	6.1	.69	4.3***	.65
L L	91	23.3	7.5	.79	19.0	7.3	.77	4.3***	.66
S H	83	23.1	6.5	.72	19.7	6.1	.67	3.4***	.79
S L	81	24.5	7.3	.82	18.0	6.6	.74	6.5***	.70
C	99	14.2	3.7	.37					

* Significant at the 5 percent level of confidence
** Significant at the 1 percent level of confidence
*** Significant at the 0.1 percent level of confidence

TABLE 5

MEANS, STANDARD DEVIATIONS, AND STANDARD ERRORS OF MEANS FOR IMMEDIATE RECALL AND DELAYED RECALL SCORES, FOR AIR FORCES POPULATION

Group	No. of cases	Immediate Recall Mean	S.D.	SE_m	Delayed Recall Mean	S.D.	SE_m	Mean Diff.	r
Total Score									
L H	107	54.7	13.1	1.27	42.7	10.7	1.04	12.0***	.63
L L	107	56.1	11.9	1.16	44.8	10.2	.99	11.3***	.61
S H	107	52.6	13.6	1.32	44.2	10.8	1.05	8.4***	.04
S L	96	50.6	13.0	1.33	42.4	9.0	.93	8.2***	.55
C	86	40.5	8.5	.92	35.8	6.4	.69	4.7***	.38
T_2 Score									
L H	107	37.5	9.0	.87	28.7	7.6	.73	8.8***	.58
L L	107	40.4	9.2	.90	30.8	7.6	.74	9.6***	.64
S H	107	37.9	10.7	1.04	30.3	7.8	.76	7.6***	.63
S L	96	35.5	9.6	.99	28.6	6.9	.71	6.9***	.46
C	86	27.2	6.5	.70	24.2	5.3	.57	3.0***	.30
V_1 Score									
L H	107	17.2	5.0	.48	14.0	4.2	.41	3.2***	.49
L L	107	15.7	4.2	.41	14.0	4.0	.39	1.7**	.36
S H	107	14.7	4.1	.40	13.9	4.1	.39	.8*	.45
S L	96	15.0	4.5	.46	13.8	3.5	.36	1.2**	.45
C	86	13.3	3.4	.36	11.6	2.7	.30	1.7***	.34
V_3 Score									
L H	107	18.9	4.7	.46	13.7	3.9	.38	5.2***	.45
L L	107	20.4	4.7	.46	14.7	4.2	.40	5.7***	.58
S H	107	19.4	5.7	.56	14.6	4.2	.41	4.8***	.65
S L	96	15.0	4.7	.48	13.6	3.3	.34	1.4***	.34
C	86	13.1	3.7	.41	11.9	3.4	.37	1.2*	.16
V_4 Score									
L H	107	18.6	5.3	.52	15.0	4.5	.44	3.6***	.52
L L	107	20.0	5.5	.54	16.0	4.7	.45	4.0***	.44
S H	107	18.5	5.9	.57	15.7	4.7	.46	2.8***	.44
S L	96	20.6	6.2	.63	15.0	4.7	.49	5.6***	.40
C	86	14.0	3.8	.41	12.3	3.3	.36	1.7***	.19

* Significant at the 5 per cent level of confidence
** Significant at the 1 per cent level of confidence
*** Significant at the 0.1 per cent level of confidence

TABLE 6

MEANS, STANDARD DEVIATIONS, AND STANDARD ERRORS OF MEANS FOR EACH TEST SCORE FOR EACH EXPERIMENTAL GROUP (COLLEGE METEOROLOGY CLASSES)

Group	N	Total Score			T_2 Score			V_1 Score			V_3 Score			V_4 Score		
		M	S.D.	SE_m	M	S.D.	SE_m	M	S.D.	SE_m	M	S.D.	SE_m	M	S.D.	SE_m
LH	33	53.4	16.2	2.87	35.8	10.9	1.93	17.7	5.8	1.02	17.4	5.4	1.02	18.4	6.0	1.05
LL	78	56.2	14.0	1.59	39.7	10.6	1.20	16.5	4.5	.51	19.2	5.7	.65	20.5	5.6	.63
SH	46	48.8	12.0	1.80	33.0	8.4	1.25	15.8	4.3	.64	15.8	4.5	.67	17.2	4.9	.73
SL	82	47.3	13.3	1.48	32.7	9.2	1.02	14.6	4.8	.53	13.8	4.9	.54	18.9	5.0	.56
C	85	41.1	13.9	1.52	27.7	9.8	1.07	13.5	4.9	.53	13.9	5.3	.58	13.8	5.2	.57

CONCLUSIONS

The findings for the study may be summarized as follows:

The more that is included in a film, the more will be learned, in absolute amount of that information (e.g., the V_1 score finding for Long Heavy group). However inferences about non-included but related information may make up for failure to present it explicitly in the film. The superiority of the Long Light or Short Heavy versions may be attributed to such inferences.

The data suggest that as more and more information is presented interferences are set up that result in less efficient learning of any particular part (e.g., the findings on the V_4 score - the Short Light Group generally did better than any other group on these items.)

Finally, it seems clear that packing more and more information into a film yields only very slight increments in total measured learning. In no case did the Long Heavy Group seem to learn anything approaching twice as much as the Short Heavy or Long Light Groups, nor did these latter learn twice as much as the Short Light Group.

Analysis of the test performance suggested that the films were rather difficult for the populations used, and observation of the attitude and performance of the groups suggested that they were not very well motivated or very interested. It should be noted that this interpretation is not inconsistent with the fact that the reading level of the scripts (as measured by the Dale-Chall formula) is at the seventh or eighth grade. These films seemed to be conceptually difficult. Furthermore, although each fact was couched in simple words, so many facts were presented per unit of time (about 4 per minute in the lightly packed versions) that grasping a large proportion of them was unlikely.