DOCUMENT RESUME

ED 057 992	24	RE 003 905

AUTHOR	Mosberg, Ludwig
TITLE	Measurement of Information Gain from Written Discourse.
INSTITUTION	Southwest Regional Educational Lab., Inglewood, Calif.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
REPORT NO	TR-29
BUREAU NO	BR-6-2865
PUB DATE	19 Oct 70
NOTE	21p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Discourse Analysis; *Grade 5; Learning Characteristics; Post Testing; *Pretesting; *Reading Comprehension; Reading Readiness; *Reading Research; Serial Ordering

ABSTRACT

A pretest/post-test procedure for measuring information gain from discourse was investigated together with several other aspects of discourse processing. The main purpose was to determine the effect of a pretest on discourse learning as measured by post-test performance. The study also investigated (1) serial position effects in learning from discourse, (2) learning of factual versus relational information, and (3) information chunking of discourse material. Four hundred fifth-graders were used as subjects. The results indicated that (1) the pretest was an essentially neutral event, neither facilitating or depressing post-test performance; (2) almost all learning or retention was on factual as opposed to relational information; (3) negative recency serial position effects were obtained as a function of the order in which the information was presented in the passage, but no serial position effect was obtained as a function of test item order; and (4) no evidence was obtained for information chunking on a supra-sentence level. Tables, figures, and references are included. (Author/AW)

OE-NCERD
RE
BR-6-2865
PA24

SOUTHWEST REGIONAL LABORATORY FOR EDUCATIONAL RESEARCH & DEVELOPMENT

Measurement of Information Gain from Written Discourse

TR29 19 OCTOBER 1970

Published by Southwest Regional Laboratory for Educational Research and Development, a public agency supported as a regional educational laboratory by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of the Office of Education, and no official endorsement by the Office of Education should be inferred

MEASUREMENT OF INFORMATION GAIN FROM WRITTEN DISCOURSE

Ludwig Mosberg[1]

ABSTRACT

A pretest-posttest procedure for measuring information gain from discourse was investigated together with several other aspects of discourse processing. The main purpose was to determine the effect of a pretest on discourse learning as measured by posttest performance. The study also investigated: 1) serial position effects in learning from discourse; 2) learning of factual versus relational information; 3) information chunking of discourse material. Four hundred fifth-graders were used as Ss. The results indicated that: 1) the pretest was an essentially neutral event, neither facilitating or depressing posttest performance; 2) almost all learning or retention was on factual as opposed to relational information; 3) negative recency serial position effects were obtained as a function of the order in which the information was presented in the passage but no serial position effect was obtained as a function of test item order; 4) no evidence was obtained for information chunking on a supra-sentence level.

[1] I wish to thank Paula Mindes, Gary Verna, and Evelyn Hatch for preparation of materials, data collection, and data analysis.

MEASUREMENT OF INFORMATION GAIN FROM WRITTEN DISCOURSE

Marks and Noll (1967), and Mosberg and Shima (1969), have defined comprehension of written discourse as the process and ability to extract, recall, and evaluate new information from a language stimulus. This definition differentiates between the measurement of the gain of new information and the measurement of new information plus whatever prior knowledge of the subject-matter of the discourse the individual may already possess. Clearly, in comprehending language, prior experience and knowledge must be brought to bear; but the end result of the comprehension process is demonstrated when something new is learned or understood.

Defining comprehension in this manner necessitates the development of dependent variables which permit the measurement of information gain. One such measure is a pretest-posttest procedure in which the S is tested on the information given in the passage prior to exposure of the passage and is then tested again subsequent to passage reading. An increase in performance from pretest to posttest is then hypothesized to represent amount of information gain. While pretest-posttest procedures are common in educational and psychological research, comprehension has not typically been measured in this manner. Rather, comprehension has been typically measured by posttest alone. Furthermore, the pretest-posttest procedure assumes that the pretest operates as a neutral event in the sense that performance on the posttest is taken to be a result of exposure to the treatment intervening between the two tests and not of the pretest per se. However, whether the difference between pretest and posttest performance is solely the result of passage reading has not been established.

It seems tenable that the pretest may not be a neutral event and, therefore, may influence subsequent passage reading and posttest performance. Several possibilities exist:

(1) The pretest items may operate as advanced organizers or cues concerning the relevant information in the passage and, in consequence, facilitate performance on the posttest (Gustafson & Toole, 1969).

(2) Conversely, it is possible that pretesting procedures result in posttest perseveration or fixation of incorrect pretest responses, thereby depressing posttest performance.

(3) Finally, the pretest may have no effect on posttest performance. In a recent study using older Ss Gustafson and Toole found, contrary to prediction, that the pretest had no appreciable effect on posttest performance when half the posttest items were used as a pretest. However, the reading passage was

extremely long and difficult (introduction to computers) and the posttest was administered one week after the pretest. Thus, it is possible that the absence of a pretest effect was due to the complexity of the material and the length of the delay between pretest and posttest.

The study reported here was designed to investigate pretest effects and two other related factors: serial position and type of information (verbatim versus substance learning).

Serial position effects. Deese and Kaufman (1957) found typical verbal learning serial position effects from discourse, i.e., recency and primacy effects. However, Rothkopf (1962) found no such effect as a function of order of information in the passage but did find a serial position effect as a function of the order of test items. The procedures and materials used in these two studies were sufficiently different to make evaluation difficult, at best. In the present study serial position effects of both order of information in the passage and order of items on the test were investigated.

Verbatim versus substance learning. A number of studies (English, Welborn & Killian, 1934; Cofer, 1941; Vernon, 1951; Yavuz, 1963; Sachs, 1967) have indicated differential effects on recall or recognition as a function of verbatim and substance learning. All but one of these studies (English et al., 1934) used either number of trials to learning or recall scores as the dependent measure. These studies show that substance recall or learning is superior to verbatim learning. English et al., however, used a recognition task and found that verbatim recognition scores were higher than substance scores. The present study attempted to shed further light on this matter using a multiple-choice recognition task. For this purpose, verbatim and substance items were written for each test passage. Verbatim items were defined as items tapping factual information contained in a single sentence. Substance items were defined as items tapping information of a relational nature, wherein the information was embedded in two or more sentences. These definitions distinguish verbatim from substance information in terms of the type of information (factual and relational) recalled or recognized whereas previous definitions (e.g., Cofer, 1941) distinguish verbatim from substance on the basis of the form of recall or recognition (word for word versus paraphrase) of essentially the same information.

Finally, the study attempted to provide preliminary data on information chunking from discourse. The question of interest was whether information is stored or retrieved in larger units than the sentence.

Method

Materials. Ten reading passages were chosen from the SRA Reading Laboratory Kits (Parker, 1963, 1964). The length of these passages ranged from 142 to 153 words with a mean length of 146 words. The Dale-Chall (1948) readability formula indicated that each passage was in the range of fifth-grade difficulty, the mean difficulty being at grade 5.6. The passages were all nonfiction content. For each passage nine four-alternative multiple-choice items were constructed. Placement of the correct alternative was counterbalanced over positions. Each passage was divided into thirds and three items were written for each third of the passage. Two types of items were constructed. Verbatim items tested recognition of specific factual information contained in a single sentence in the passage. Substance to correctly answer the item was given in at least two sentences and the S was required to make a logical inference from the facts or to understand and recognize the relationship between facts. All items were written using vocabulary found in the passage. Since there was an odd number of items, the tests for five randomly selected passages contained five verbatim and four substance items while the remaining five passages had tests with four verbatim and five substance items.

Four item orders were used for testing. In Order 1 the items were ordered in the same sequence as the information was presented in the passage. The remaining three orders were obtained in the following manner: Order 1 for each of the 10 passages was divided into three subsets, the first three items, second three items, and the last three items. The three subsets were then sequenced according to the Latin square design (Table 1). Within each subset of three items one randomized order of items was used for each of the 10 passages.

TABLE 1

TEST ITEM ORDERS FOR EACH SUBSET OF THREE ITEMS

Order 1	1	2	3
Order 2	1	3	2
Order 3	2	1	3
Order 4	3	2	1

Procedure. Subjects were tested in groups of approximately 20 to 25 per group. Each S was randomly assigned to a testing group and each group was assigned to an experimental condition.

The paradigms for the experimental conditions are shown in Table 2. Group A received the comprehension test items as a pretest followed by the test passage. Upon completion of the reading, the same comprehension items were administered. For half the Ss the items on the posttest were given in the same order that they appeared on the pretest and for the other half the items were in one of three different orders on the posttest. The passage was not available to the Ss at the time of testing. Immediately following the posttest an unrelated task (arithmetic problems) was administered for 5 minutes. A second posttest was then administered. Again, for half the Ss the items were presented in the same order as on the previous test and for the other half the order was changed.

Group B differed from Group A only in that the immediate posttest was not given. Group C differed from Group A only in that no pretest event occurred for this group. Group D was treated identically to Group A except that the group's "pretest event" was an unrelated test (a test appropriate to some other passage than the one presented) for Group D1 or an unrelated task (arithmetic problems) for Group D2. Group E was treated identically to Group A except that the group's "training event" was an unrelated passage for Group E1 or an unrelated task (arithmetic problems) for Group E2. Group F was identical to Group A through the immediate posttest except that the pretest consisted of only half (4) of the posttest items. Group G was identical to Group B except that the pretest consisted of only half of the posttest items.

Each treatment component was placed in a separate envelope lettered from A to E. Each S received a set of four or five envelopes, depending upon his assigned condition. The Ss were instructed not to open any envelope until the E instructed him to. At the end of each treatment event E told the Ss to return the material to the appropriate envelope and to open the next one, which E referred to by letter name. Prior to each event, E described the tasks to be negotiated and asked Ss to do as well as they could on each task.

Initial pilot work indicated that over 95% of the Ss would complete both the pretest and posttests in less than 4.5 minutes and read the passage in at most 2 minutes. The arithmetic problems were so designed that the Ss could adequately negotiate each problem but could not complete the entire task in less than 5 minutes.

Subjects. During an initial phase of the study 40 Ss were assigned to Groups A through E. A preliminary analysis of the data suggested a need to broaden the study (to include Groups F and G) and to replicate findings (for Groups A, B, and C). During the terminal phase of the study, 40 Ss were assigned to Groups F and G and 40 additional Ss to Groups A, B, and C. Procedures for original and replication Groups A, B, and C were identical.

TABLE 2

TREATMENT GROUPS

Group	Number Tested	Pretest Event	Training Event	Immediate Posttest	5-Min Delay, Unrelated Task	Delayed Posttest
A	40+40	Full test	Passage	X	X	X
B	40+40	Full test	Passage		X	X
C	40+40		Passage	X	X	X
D1	20	Unrelated test	Passage	X	X	X
D2	20	Unrelated task	Passage	X	X	X
E1	20	Full test	Unrelated passage	X	X	X
E2	20	Full test	Unrelated task	X	X	X
F	40	Half test	Passage	X		
G	40	Half test	Passage		X	X

In all, 400 fifth-graders from Southern California schools served as Ss, with assignments to groups as indicated in Table 2. California Reading Achievement Test scores were obtained from school records and were analyzed to determine whether the various Groups were comparable in reading achievement. No statistically significant differences were found either between Groups A through F or between the original and replicated Groups A, B, and C. The mean grade-level reading score across groups was 5.0.

Pre- and posttest performance. The mean proportions of items correct on pretest and posttests are shown in Table 3. For Groups A, B, and C the means for each replication are given both separately and combined. Performance on the pretest (T_1) was consistently above chance for all groups receiving a pretest. An analysis of variance comparing pretest performance across groups showed no significant difference ($p < .05$) on T_1. In addition, no reliable differences were obtained on T_1 performance between original and replications in Groups A, B, and C. Since no reliable T_1 differences were obtained, direct comparisons of posttest scores were made. The data for posttest performance was first analyzed excluding the replication data. Analysis of variance comparing Subgroups D1 and D2 and Subgroups E1 and E2 yielded no reliable differences; consequently, the data for D1 and D2 and E1 and E2 were pooled for purposes of all further analyses.

An analysis of variance of the proportion correct for T_2 indicated a significant difference among groups $F(4,195) = 4.46$, $p < .01$. A Duncan Multiple Range Test indicated that both Groups B and C scored reliably higher on T_2 than Groups A, D, and E. No other comparisons differed reliably. An additional analysis of variance between T_1 and T_2 scores for Groups A, B, and E indicated significantly higher scores on T_2 than on T_1, $F(1,117) = 26.47$, $p < .01$. Comparing T_1 and T_2 scores for Groups A and E and comparing T_2 performance of Groups A, B, and C, it appeared that the pretest-immediate-posttest procedure for measuring information gain was inadequate. That is, Group A was comparable to Group E, which never received the relevant reading passage, while Group B which received a delayed posttest and Group C which received no pretest were reliably superior on posttest performance to Group A. These results, however, were suspect particularly in light of the fact that Group D, which received either an irrelevant pretest or an unrelated task (arithmetic problems), should have shown results more comparable to Group C. This was clearly not the case. The results were further suspect in that the data were contrary to the results of Gustafson and Toole discussed earlier. In consequence it was decided to conduct a replication for Groups A, B, and C and to add Groups F and G as a check on these findings. The results of the replication are given in Table 3. Analysis of variance indicated no reliable differences among groups. In addition, when original and replication results were

TABLE 3

MEAN PROPORTION CORRECT ON PRE- AND POSTTESTS

Group	Pretest (T_1)			First Posttest (T_2)			Second Posttest (T_3)
	Replications		Combined	Replications		Combined	
	R_1	R_2		R_1	R_2		
A	.42	.42	.42	.49	.58	.53	.51
B	.45	.37	.43	.58	.45	.52	---
C	---	---	---	.62	.53	.58	.61
D1	---	---	---	.51	---	---	.55
D2	---	---	---	.47	---	---	.51
E1	.46	---	---	.50	---	---	.51
E2	.35	---	---	.42	---	---	.44
F	---	.38	---	---	.52	---	---
G	---	.36	---	---	.48	---	---

combined the findings based on the original groups were overturned in that no reliable differences were obtained. Moreover, when T_2 replication results for Group C were compared with the original results of Group D, performance was comparable, as would be expected. The results of Groups F and G further indicate that results for the original and replication groups combined are a more accurate estimate of the true differences than the results indicated by the data of the original groups. The mean proportions correct on the posttest for items presented to Groups F and G on the pretest and for items not presented on the pretest are given in Table 4. It is clear that Ss in the two groups perform equally well on the posttest regardless of whether the item was presented on the pretest. These data confirm the earlier results reported by Gustafson and Toole.

TABLE 4

PROPORTION CORRECT ON POSTTEST FOR ITEMS PRESENTED ON PRETEST AND ITEMS NOT PRESENTED ON PRETEST

Group	Items Presented on Pretest	Items Not Presented on Pretest
F	.51	.51
G	.46	.49

The previous discussion has been concerned only with the results for T_1 and T_2 tests. The results of T_3 provide no further information on treatment effects. The results of T_3, presented in Table 3, show essentially no forgetting over the 5 minute delay between T_2 and T_3.

The determinants of the differences between the original and replication results are not immediately obvious. The Subject samples appear comparable in that both samples had identical mean reading achievement scores; the variance in reading scores were comparable for the two samples; the samples were drawn from similar socioeconomic areas in Southern California and no other evidence could be found to suggest that Ss differed in any significant way.

However, the geographical settings of the replications did differ. A second possibility is differences in experimental procedures. All experimental procedures were ostensibly replicated with two exceptions: 1) The original data were collected by a female E while the replication data were collected by a male E. 2) The original data were collected in the spring while the replication data were collected in the fall.

A third possibility is that the difference in results was due to the unreliability of the test instruments or in scoring and analysis. Scoring and analyses were checked and double checked; no significant errors were found. A test-retest reliability coefficient was computed on the data of Group E which was the only group that did not receive the relevant passage between pretest and posttest. However, since there were only 40 Ss in this group and since there were 10 different tests with only four Ss receiving each test, any reliability coefficient would be spuriously low. In spite of these deficiencies, the test-retest reliability coefficient was .70. While lower than what is generally considered adequate, the coefficient appears acceptable considering the number of Ss on which it is based, the confounding due to collapsing over 10 different tests, and test length (only nine items). The conclusion drawn from the reliability check is that while unreliability of the tests cannot be rejected as a possible explanation for the difference in the original and replication results, this explanation is quite unlikely.

Finally, it is possible that a Type I error occurred in the original results or that a Type II error occurred on the replication. For reasons discussed previously and in lieu of any other satisfactory evidence it is assumed that a Type I error did, in fact, occur in the original results. Briefly, this conclusion is based on the results for Groups F and G; comparability of Group D with the replication of Group C; the similarity of the replication results with the findings of Gustafson and Toole.

Performance on verbatim versus substance items. Half of the 10 passage tests consisted of four verbatim and five substance items while the other half consisted of five verbatim and four substance items. Verbatim items tapped information given in a single sentence using the original sentence vocabulary wherever possible. The substance items tapped information which required the S to combine information from two sentences. The two sentences were not necessarily adjacent in the passage. Pretest and posttest performance by item types is of considerable interest. Table 5 presents the mean proportion of correct responses for verbatim and substance items as a function of group and test trial. The means for the three replicated groups are shown in the bottom half of Table 5. Since there were only two of each type of item on the pretest of Groups F and G, the data for these groups are not included in this analysis.

Analysis of variance indicated that on T_1 there were no significant between-groups or between-replication differences in correct responding to verbatim and substance items; nor were there reliable interaction effects. On the first posttest (T_2) there was a reliable difference between item types on both replications, $F(1,195) = 11.19$ and $F(1,117) = 16.09$, $p < .01$, respectively. Subjects across groups performed significantly better on verbatim items than on substance items. This was true for all groups except Group E for which no differences would be expected since this group did not see the passage to which the tests referenced. Since the difference in performance on item types was not significantly different on T_1, the conclusion to be drawn from these data is that most of the learning or retention measured on T_2 was for

TABLE 5

MEAN PROPORTION CORRECT FOR VERBATIM AND SUBSTANCE ITEMS AS A FUNCTION OF CONDITION AND TEST TRIAL

Group	Verbatim			Substance		
	(T_1	T_2	T_3)	(T_1	T_2	T_3)
A	.43	.57	.54	.41	.38	.43
B	.46	.59	---	.47	.56	---
C	---	.53	.54	---	.43	.50
D	.38	.44	.48	.42	.45	.47
E	---	.67	.61	---	.56	.59
Mean	.42	.56	.54	.43	.47	.49
		Replication				
A	.41	.63		.44	.53	
B	.39	.50		.36	.39	
C	---	.61		---	.45	
Mean	.40	.58		.40	.46	

verbatim information. This makes intuitive sense since the verbatim items tapped simple factual information while the substance items tapped the relationship between various facts and thus represent more complex information. It is also interesting that for Groups A and C there is a slight drop in verbatim performance on T_3 and a slight increase in substance performance on T_3. While the differences were not statistically significant, they suggest the possibility that whatever forgetting takes place over time will affect specific factual knowledge rather than substantive information.

Previous studies of verbatim versus substance learning indicate that substance learning is superior to verbatim learning when the dependent variable is trials to criteria or free recall (Cofer, 1941; Yavus, 1963; Sachs, 1967). The one study using a recognition measure (English et al., 1934) found performance on verbatim items better than on substance items. The results of the present study support the English results and suggest a differential effect as a function of the dependent variable.

Test item order and serial position effect. Four test item orders were used. One order sequenced the items in approximately the same order in which the information occurred in the passage; the other three orders counterbalanced item order (see Table 1). Half the Ss received the same item order on all tests while the other half received a different order on each test.

The first question is whether receiving the test items in the same order on pretest and posttest facilitates performance. Using analysis of variance, no reliable difference was obtained as a function of same or different order over pretest and posttest or between first and second posttest.

It was originally suspected that Ss who receive the test items in the same order as the information in the passage is ordered would perform better on the posttest since the information might be stored serially in memory. An analysis, therefore, was done to determine whether there were any differential effects of test order per se. To avoid confounding with order of items on the pretest, only those Ss who got the same order on all tests were used in this analysis. Table 6 presents the mean proportion correct for the two posttests by groups and collapsed over groups. Group E was not included in this analysis. Only the data of the original groups were used in the analysis.

TABLE 6

MEAN PROPORTION CORRECT FOR EACH OF FOUR TEST ITEM ORDERS BY GROUPS

Group	Normal Order Order 1		Order 2		Order 3		Order 4	
	T_2	T_3	T_2	T_3	T_2	T_3	T_2	T_3
A	.58	.60	.42	.49	.62	.56	.60	.55
B	.60	---	.64	---	.60	---	.53	---
C	.53	.58	.51	.58	.58	.62	.45	.49
D	.69	.59	.62	.65	.62	.65	.42	.42
Overall	.60	.59	.55	.57	.61	.61	.50	.49

The data show no consistent relationship between order and group. The analysis of variance yielded no reliable differences among orders except that performance under Order 4 was reliably poorer than that for any of the other three orders. Order 4 represents the greatest amount of change from input order (Order 1), as can be seen in Table 1. Thus, the significantly poorer performance under Order 4 may suggest that a relatively large shift in input-output order results in poorer performance and that this variable cannot be ignored in testing information gain.

One of the purposes of this study was to shed further light on the serial position effect in learning from discourse. Deese and Kaufman (1957) reported data which shows a typical serial position curve, i.e., both primacy and recency effects. Rothkopf (1962) in a later study was unable to replicate this effect in terms of the order in which the information was given in the passage. He did, however, find a serial position effect as a function of the order of test items. Since in this present study there were four orders of test items (one of which corresponded to the order of information in the passage) it was possible to assess the effects of both serial position as a function of order of information in the passage and serial position as a function of test-item order. To determine the serial position effect as a function of order of information in the passage, the test items were rearranged to coincide with the order of information (i.e., Order 1). For the serial position effect as a function of test item order the items were left in the order in which they occurred on the test. Separate analyses were performed for T_2 and T_3. These data were also separately analyzed for the replication. To reduce the variability from one position to the next, the nine items were divided into blocks of three. Figure 1 presents the mean proportion correct by item block for test item order and for order of information. It can readily be seen that serial position of test items has no effect on performance. When the items are rearranged to correspond to the order in which the information was presented however, there is a negative recency effect, that is, there is a drop in performance on the last item block. A Treatment X Subjects analysis of variance indicated a reliable item block effect, $F(2,318) = 11.82$, $p < .01$. Figure 2 presents the same curves for T_3. It will clearly be seen that the curve is almost identical to Figure 1. Similarly, the data of the three replicated groups are plotted in Figure 3 and the curve is almost identical to those in Figures 1 and 2. Again, an analysis of variance indicated a reliable item block effect, $F(2,234) = 3.22$, $p < .01$.

These results support neither Deese and Kaufman nor Rothkopf. The results show neither the typical verbal learning serial position effect found by Deese and Kaufman nor do they show a serial position effect of test item order as found by Rothkopf. In this study, shift in input-output packaging of information in discourse appeared to be more consequential than serial position effect per se.

Chunking of information in learning from discourse. The question of interest here concerns how information in discourse is stored and retrieved. That is, if information is stored in larger units than the word, phrase, or sentence, then in this experiment, the conditional probabilities of getting any two items correct or incorrect should depend on the temporal or spatial proximity of the information in the passage tested by the two items. For example, if chunking of information occurs across sentences and if a S responds correctly to Item 1, which tests for information contained in Sentence 1 of the passage, then the probability of a correct response to Item 2, which taps information given in Sentence 2, should be higher than the conditional probability of a correct response to Item 5, which taps information given in Sentence 5. Four analyses were done to test this hypothesis. In the first, the proportion of correct responses to both items for each combination of two items was computed. The second analysis involved the proportion of incorrect responses to both items for each combination. The third and fourth analyses evaluated the conditional probabilities of a correct response on the first item and an incorrect response on the second, and the conditional probabilities of an incorrect response on the first item and a correct response on the second. These proportions, for all four analyses, were then subtracted from the cross-products of their corresponding independent probabilities of each item of the pair. This was done to correct for serial position effects. It was expected that the closer two items were in terms of the order of information, the greater the difference between the conditional probabilities and the cross-products of the independent probabilities. The results of these analyses indicate no differences either

Fig. 1. Mean Proportion Correct Per Item Block (1-3 4-6 7-9) For T_2 - Groups A,B,C, and E

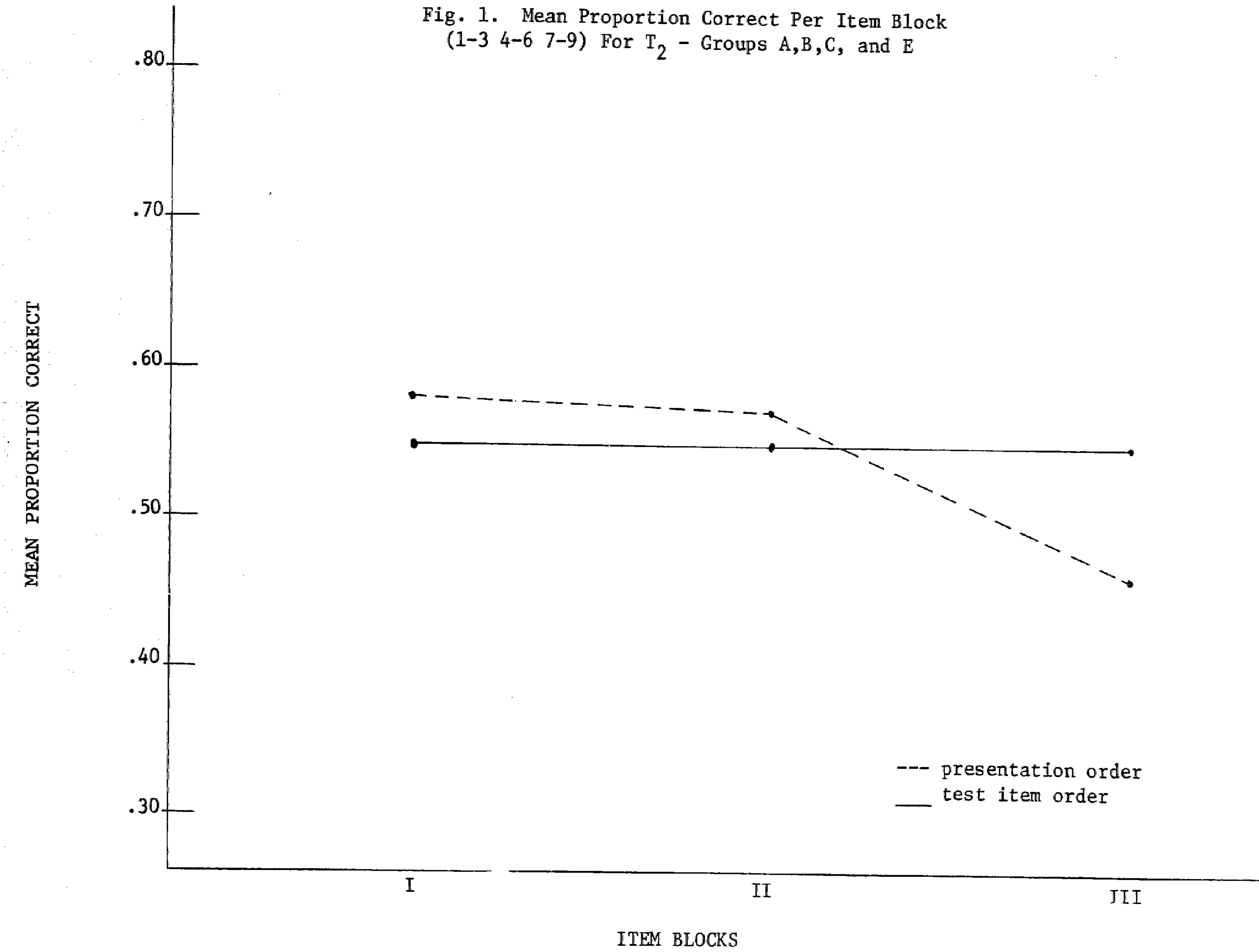


Fig. 2. Mean Proportion Correct Per Item Block
(1-3 4-6 7-9) For T_3 - Groups A,C, and E

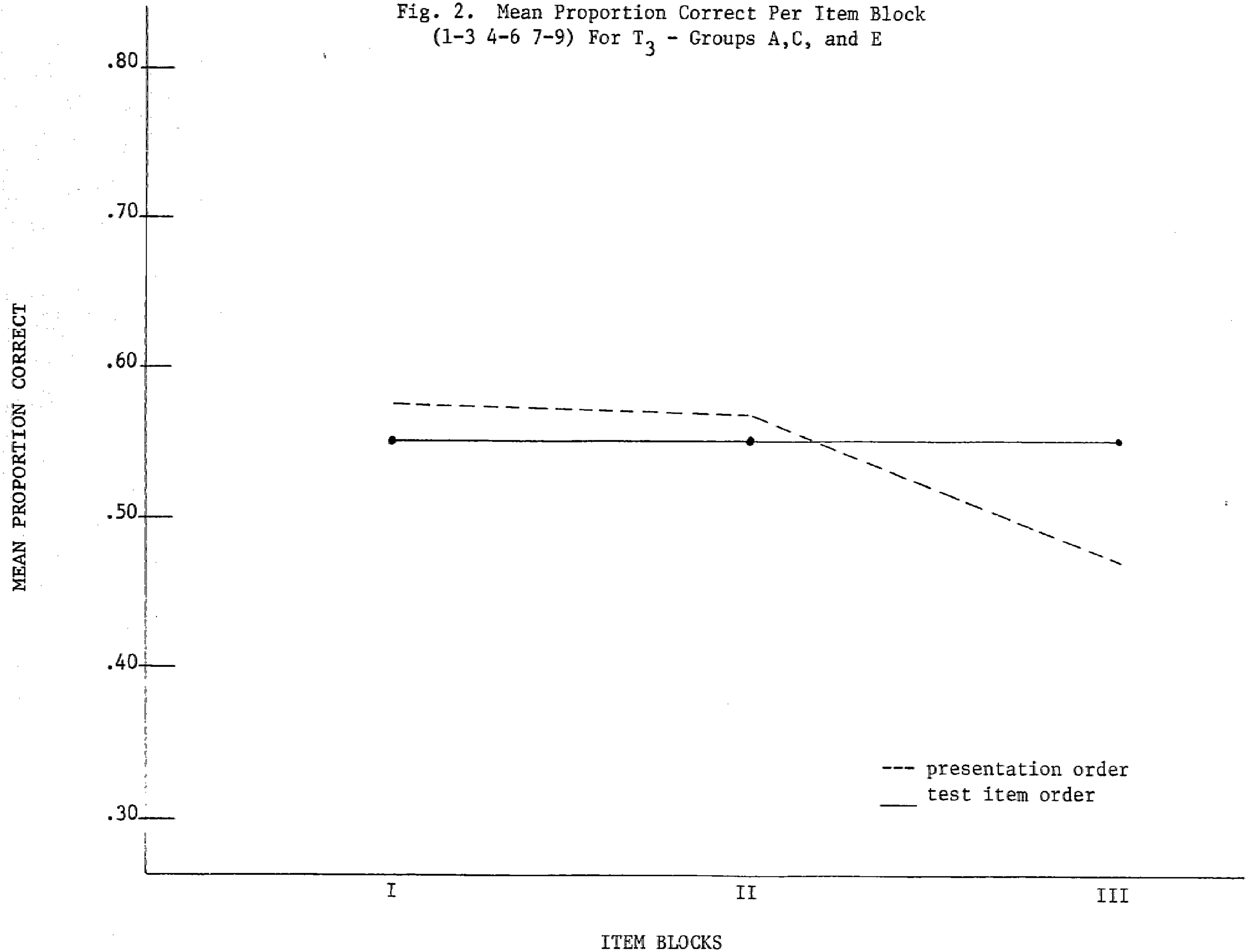
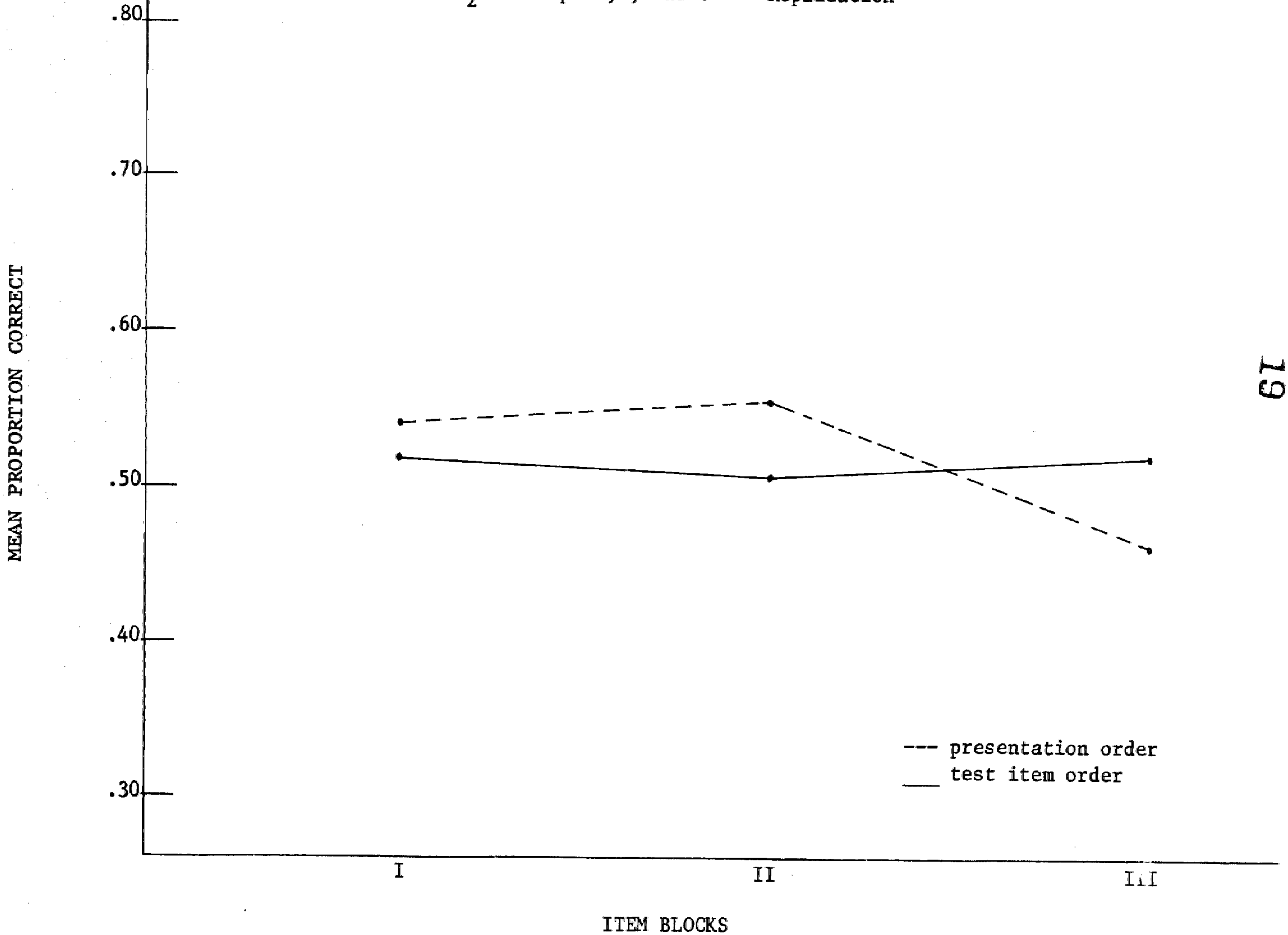


Fig. 3. Mean Proportion Correct Per Item Block (1-3 4-6 7-9) for T_2 - Groups A,B, and C --- Replication



within or across groups in any of the four analyses. Items appeared to be completely independent of each other. While this test of information chunking was quite gross the results were nevertheless disappointing.

Conclusions

The results indicate that while there is a good deal of variance in pretesting effect, the overall effect of pretesting is substantially neutral. At least for the fifth-grade population of children whose reading achievement is within the normal range, the pretest-posttest procedure can be reasonably applied as a measure of information gain.

The verbatim versus substance item performance comparisons are provocative. A more detailed analysis of types of information and items which tap this information appears warranted. Further, methods for training children to orient to substance information would appear to be a profitable line of investigation.

Similarly the test item order and serial position effects suggest a careful study of: 1) input-output task characteristics, and 2) organizational and structural properties of discourse with the view of optimizing such properties for information gain.

REFERENCES

Cofer, C.N. A comparison of logical and verbatim learning of prose passages of different length. American Journal of Psychology, 1941, 54, 1-20.

Dale, E., & Chall, J. A formula for predicting readability. Educational Research Bulletin, 1948, 27, 11-28.

Deese, J., & Kaufman, R.A. Serial effects in recall of unorganized and sequentially organized verbal material. Journal of Experimental Psychology, 1957, 54, 180-187.

English, H.B., Welborn, E.L., & Killian C.D. Studies in substance memorization. Journal of General Psychology, 1934, 11, 233-260.

Gustafson, H.W., & Toole, D.L. Effects of adjunct questions, pretesting and degree of student supervision on learning from an instructional text. Paper presented at American Educational Research Association Convention, Los Angeles, 1969.

Marks, E., & Noll, G.A. Procedures and criteria for evaluating reading and listening comprehension tests. Educational and Psychological Measurement, 1967, 27, 335-348.

Mosberg, L., & Shima, F. Comprehension of Connected Discourse. Technical Report No. 12, 1969, Southwest Regional Laboratory, Inglewood, California.

Parker, D.H. SRA Reading Laboratory: llb. Chicago: Science Research Associates, 1963.

Parker, D.H. SRA Reading Laboratory: llla. Chicago: Science Research Associates, 1964.

Rothkopf, E.Z. Learning from written sentences: Effects of order of presentation on retention. Psychological Reports, 1962, 10, 667-674.

Sachs, J.A. Recognition memory for syntactic and semantic aspects of connected discourse. Perception and Psychophysics, 1967, 2, 437-442.

Vernon, M.D. Learning and understanding. Quarterly Journal of Experimental Psychology, 1951, 3, 19-23.

Yavuz, H.S. The retention of incidentally learned connotative responses. Journal of Psychology, 1963, 55, 409-418.