DOCUMENT RESUME

ED 090 302 TM 003 588

AUTHOR Dyer, Jean L.
TITLE The Effect of Memory and Higher Level Cognitive Tests Upon Learning.
INSTITUTION Louisville Univ., Ky.
SPONS AGENCY Louisville Univ., Ky. School of Education.
PUB DATE Apr 74
NOTE 31p.; Paper presented at American Educational Research Association Annual Meeting (Chicago, Illinois, April 15-19, 1974)

EDRS PRICE MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS Ability; Cognitive Processes; *Cognitive Tests; High School Students; *Learning; *Memory; *Response Style (Tests); *Sex Differences; Social Studies; Test Wiseness

ABSTRACT

Repeated exposure to tests composed of memory items was compared to repeated exposure to tests including higher-level cognitive items and to tests composed of both item types. Other design factors were sex, ability level, subject matter sequence (science and social studies), and learning session. (A sample of 2008 eleventh-grade students participated in phase one of the study and a total of 288 participated in the final study). Although significant main effects for the test conditions did not occur, treatment by sex interactions on both memory and higher-level tests indicated that male performance was enhanced by a variety of item types, female performance by memory items. (Author/MLP)

The Effect of Memory and Higher-Level Cognitive Tests Upon Learning

Jean L. Dyer
School of Education
University of Louisville

Paper presented at the annual meeting of the American Educational Research Association, Chicago, Illinois, April 15-19, 1974.
Session 21.21

The Effect of Memory and Higher-Level Cognitive Tests Upon Learning[1]

Jean L. Dyer
University of Louisville

Both elementary and secondary students are constantly exposed to tests and classroom questions emphasizing memory (Billings, 1971; Davis & Hunkins, 1966; Gall, 1970). However, little is known about the effects of such repeated exposure, particularly that of tests, upon students' problem solving ability.

Does repeated exposure to a particular type of test produce a psychological set for that type of test? If a "psychological set" (Luchins & Luchins, 1959) to process information appropriate for a particular type of test item (e.g., memory) can be established by the continuous use of tests composed of questions primarily of that type, then cognitive processing at other levels might be inhibited (e.g., comprehension, application). Both the facilitating and inhibiting effects of set have been documented in experimental settings (Luchins & Luchins, 1959). However, Jeffrey (1969) also showed that performance on the Luchins' water jar problems correlated highly with performance on a math quiz where a switch in set was necessary to solve the remaining problems. In fact, the majority of the high school students (80%, n=30) were unable to change sets on both the math and water jar problems.

---

[1] This research was supported by the University of Louisville School of Education's Research Committee. Appreciation is expressed to the principals, counselors, and teachers who helped arrange the data collection, and to the graduate students who administered the materials. Special acknowledgement is accorded to Jennie Gehlbach and Sarah Knight who helped in the development of the passages and tests.

In the present study the effects of repeated exposure to tests composed of Knowledge items were compared to those composed of Comprehension, Application, and Analysis items (Bloom's Taxonomy, 1956). Exposure per se rather than instructions to study for a certain type of test was the focus of the study. Sets can occur without individuals being consciously aware of the fact that they are using a certain procedure or process to solve problems (Johnson, 1955), although maximum efficiency will only occur with conscious thought. The task-induced set from exposure to different item types would probably require the subject to attend to rather complex stimulus materials. Under a memory condition the material would be memorized. Under a condition with higher-level items, the individual would concentrate on understanding basic concepts, interrelating concepts, examining the underlying intent of the author, etc. This type of behavior could also be classified as mathemagenic (Rothkopf, 1970) since the cognitive level of the test item would probably "determine the nature of the effective stimuli in experimental or instructional situations (p.326)."

A few studies have examined the effect of repeated exposure to questions at different levels of cognitive functioning. In examining the role of questions inserted within prose, Watts and Anderson (1971) found that questions requiring students to apply principles to new examples produced generally superior performance on post tests than questions requiring recall of examples given in the text. The application questions enhanced performance at both the application and memory levels. Similarly, Hunkins (1968) varied the cognitive level of questions within prose by comparing memory questions to a combination of analysis and evaluation questions on a post test which included questions at each level

of Bloom's Taxonomy. Differences in favor of the analysis and evaluation questions occurred only on application and evaluation items.

A study which specifically examined the effect of examination items was that by Cooper (1967). He compared the effects of two treatment conditions; quizzes composed of knowledge (memory) items and quizzes composed of items above the memory level (Bloom's Taxonomy). The type of test did not influence student performance on a final exam which contained questions at all cognitive levels. However, McKenzie (1972) found that quizzes composed of inference items faciliated performance on a criterion inference test over the same content compared to a quiz treatment composed of factual items.

Several variables which could possibly affect the establishment of sets by classroom tests have been found in the traditional laboratory studies. Females seem to be more easily affected than males (Luchins & Luchins, 1959). Van de Greer (as reported by Ray,1967) also found that girls took on set more easily than boys and had greater difficulty in overcoming set. Intelligence has generally been found to correlate negatively, although at a low level, with set (Luchins & Luchins, 1959) with more intelligent individuals being less rigid. In general, as the number of training problems increases, the strength of set increases (Ray, 1967). This relationship has often been a negatively accelerated function of the number of training trials, with a plateau being reached after 6 to 8 trials in the typical water jar experiment (Luchins & Luchins, 1959). Ray (1967) has postulated that the increase in set strength with increasing trials may be due to self-reinforcement, since in most experimental settings the subject usually recognizes a solution as correct after has has found it.

In applied settings the variables of sex and intelligence can usually be controlled. Cooper (1967) covaried intellectual ability and males and females were randomly assigned to treatments, although sex was not included as a factor in the analysis. McKenzie (1972) randomly assigned subjects to treatments, and checked for equivalence of ability; sex was not analyzed as a design factor. Controlling for "number of trials" is more difficult, since the number of trials necessary to establish a set in the test situation has not been determined. The number of tests, the number of items, and the spacing of the tests are all relevant parameters, and probably could be varied in different amounts to produce similar effects. Cooper used a series of eight daily quizzes, each composed of ten items. McKenzie had a series of eight weekly quizzes, each composed of five items. Thus Cooper employed a more concentrated treatment than McKenzie. However, neither researcher examined the performance of Ss on the series of quizzes to determine if a set was being established, e.g., improvement on memory items as the number of quizzes increased. In both studies some type of feedback was given to the students. Because of the degree of experimental control, students were not allowed to study their performance for a long period of time, nor were they allowed to discuss their results with the teacher or other students. Thus the "self-reinforcement" phenomenon postulated by Ray (1967) could have occurred, although the reinforcement was not immediate, and the reason for the correct solution was probably not always clear to all students.

The present study compared the effects of exposure to three types of quizzes: only Knowledge items (called Memory); both Comprehension, Application, and Analysis items (called Higher); and all four item types (called Combination). In addition, the factors of sex, ability, and content area were included ---

the ability factor to increase the precision of the design, and both social studies and science materials were used to determine the generalizability of the results. Feedback was given to each subject after he had completed a test. It was expected that students would perform best on the type of test items to which they had been repeatedly exposed, this effect being strongest for females.

A pilot study was conducted to select passages and items for the final study. The reliability of item classification at the levels of Bloom's Taxonomy was determined, as well as the reliability of the test, the difficulty and discrimination levels of the items, and the students' interest in and prior knowledge of the material in the passages.

Pilot Study

Passages and Tests

Eight passages within both the science and social studies areas, approximately ten to twelve pages in length, were constructed.[2] The science passages were entitled: Atmosphere and Life; The Atom; Living Cells; Mars and Venus; The Nervous System; The Skeleton; Sound; and Spiders. The social studies passages were: Buying, Renting, and Selling; Comparative Economic Systems; Crime and Justice for Adults and Juveniles; Feeding the World's Population; Kinship Relationships in Anthropology; Law in Primitive Societies; Money in Early American History; and Political System of the USSR.

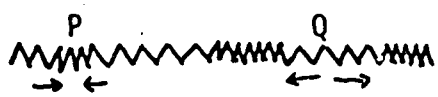
For each passage 20 Knowledge (K) and 20 Comprehension, Application and Analysis (CAA) items were written.[2] All items were multiple-choice with four alternatives. To determine the reliability of this item classification, the items were independently classified by an advanced graduate student majoring in measurement.[3]

[2] Jennie Gehlbach constructed all passages and tests except for "Money in Early American History" which was based upon Cooper's (1967) materials.

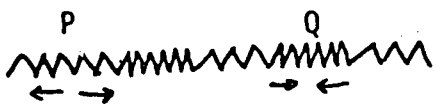
[3] Sarah Knight classified the items.

Sample sections from the passage on Sound are given below with related test items which illustrate the four levels of the Taxonomy hierarchy. A Knowledge item is presented first with the corresponding passage.

> Another kind of wave which travels in a slightly different way is a longitudinal wave. In this kind of wave, instead of moving up and down or from side to side across the direction of the movement of the wave itself, the individual parts of the vibrating object move back and forth across the same spot but in the same general direction as the wave's movement. A good example of this type of wave occurs when a long spiral spring is held horizontally and suddenly pushed together (compressed) at one end, and then pulled apart (rarefied, the opposite of compressed). If this is repeated a few times, alternate compressions and rarefactions pass down the spring away from the end being pushed and pulled. Any single point along the spring is alternately compressed and rarefied as the waves pass by.



First Instant



A Few Moments Later

> A Longitudinal Wave. All points in areas such as P and Q are alternately pushed together and pulled apart along the same general direction as the wave moves in.

Knowledge Item:

> In a longitudinal wave in a spring there are alternate large and small amounts of:
> a. compression at each point of the spring*
> b. movement of each point on the spring up and down
> c. wave frequency
> d. compression along the whole spring

A Comprehension Item (requires the ability to translate from verbal to numerical terms):

If you multiply together the frequency and wavelength of a wave, you get the speed of the wave. If a sound wave is travelling at 340 yards per second, and one complete vibration takes 1/680th of a second, what is the wavelength of the sound?

a. 1/2 yard*
b. .2 yard
c. 21 yards
d. 4 yards

Below is another passage selection with an Application item based upon a principle described in the passage, but the appropriate principle is not supplied in the item for the student as was the case with the preceding Comprehension item.

It a wave is initially given more energy, the wave is stronger or more intense. In sound waves this results in louder sounds. What in fact happens is that in a transverse wave the peaks and troughs become deeper. In a longitudinal wave the compressions and rarefactions are very greatly contrasted; the individual sections are pushed very strongly toward each other and highly compressed, or pulled apart very strongly and highly rarefied.

Application Item:

The more air is compressed, the more it heats up; the more air is rarefied, the more it cools down. Which of the following sounds could be expected to cause the biggest differences in temperature of the air particles as a sound wave passes through them?

a. a loud sound*
b. a quiet sound
c. a high sound
d. a low sound

Analysis Item (requires the student to examine basic assumptions underlying a statement):

"Explosions occurring in the stars can be seen but not heard on earth." Which of the following statements is consistent with and serves as an explanation for this fact?

a. light travels faster than sound
b. intensity of sound diminishes with time
c. the intensity of sound diminishes with the distance travelled
d. sound waves cannot travel through empty space*

Subjects and Procedure

The sample consisted of 2008 eleventh grade students from eighteen high schools. Although ability tests were not given, it seemed safe to assume that a wide range of ability was obtained since a) the schools were located in different socio-economic areas, b) in some schools the total eleventh grade population participated, and c) in some schools classes were selected for their wide range of ability.

The students were randomly divided into two groups: one which read the passages before taking the test (PASS) and the other which took the test prior to reading the passages (TEST), allowing a direct test of the amount of prior knowledge possessed by the students. The TEST group first took two tests (one science and one social studies) and was then allowed to read the corresponding passages. Each subject in the PASS group read only one passage and then took the test on it.

Packets of materials containing the passages and tests were created for each S in the PASS and TEST groups. In order to evenly distribute the passages to the students, within each group each series of sixteen passages was ordered by random permutations of sixteen numbers. Since more subjects were required for PASS than TEST, the PASS and TEST packets were then ordered in approximately a three to one ratio, and distributed to students accordingly. Of the 2008 students, 1459 were in the PASS group and 549 in the TEST group. Only 1822 Ss (91%) indicated their sex. Within PASS there were 927 males and 895 females; within TEST there were 241 males and 248 females.

Students were instructed that they would be working with materials different from their neighbor, that they would not be graded on their performance, and that they could work at their own rate. Most students finished within 50 minutes. After completing the material, PASS Ss were asked to

indicate their prior knowledge of and interest in the passage they had read (Have you read anything on this topic before? and Would you want to read more on the same topic?).

Results

On the basis of item analyses of the test and student interest in and prior knowledge of the passage content, four science and four social studies passages/tests were selected for the final study. The science passages were: Atmosphere; Mars and Venus; Nervous System; and Sound. The social studies passages were: Feeding the World's Population; Kinship in Anthropology; Money; and USSR Political System. Table 1 presents these data for each of the selected passages. For comparative purposes averages are presented for the passages which were not selected. On the average the more difficult passages were selected, so that students would be less likely to reach ceiling on the tests in the final study. In general students indicated more familiarity with the science than with the social studies topics, although test performance was similar. The selected science passages were of a higher interest level than the non-selected passages, although the interest levels of the selected and non-selected social studies passages were similar. The reliability estimates on the selected and non-selected social studies passages were similar, although for the science passages the average reliability for the non-selected passages was slightly higher than the selected. Inter-rater agreement on item classification was determined by comparing the rater's classification with the researcher's classification (each item was classified as a K or CAA item), and was expressed in terms of percentage of agreement. In general, the agreement for science items was higher than that for social studies items; and items on the selected passages had higher agreement than those on the non-selected passages. Any discrepancies in the K and CAA

Table 1

Data on PASS and TEST Groups for Each Selected Passage/Test and for All Non-selected Passages/Tests: Item Analysis, Interrater Agreement on Item Classification, and Student Reaction to Passages

Passage/Test	N		Difficulty		Discrimination[a]		KR-20		Interest	Familiar	Rater Agreement
	PASS	TEST	PASS	TEST	PASS	TEST	PASS	TEST	PASS %	PASS %	
Science-Selected											
Atmosphere	94	69	.45	.37	.38	.38	.65	.53	34.6	53.9	96.6
Mars & Venus	91	68	.45	.34	.43	.42	.69	.46	60.2	52.5	92.9
Nervous System	90	71	.48	.37	.37	.35	.69	.57	45.3	74.4	93.1
Sound	89	66	.46	.37	.33	.41	.60	.51	34.3	53.5	84.6
Mean											
Selected	91.0	68.5	.46	.36	.38	.39	.63	.52	43.6	59.1	91.8
Non-selected	93.7	67.7	.51	.39	.41	.36	.73	.41	29.8	70.0	88.6
Social Studies-Selected											
Feed Population	90	68	.52	.41	.36	.35	.69	.60	64.2	37.9	92.9
Kinship	88	66	.48	.40	.43	.26	.75	.21	33.3	17.2	78.6
Money	89	70	.42	.32	.36	.39	.53	.38	27.3	23.0	96.3
USSR Political	89	72	.39	.31	.44	.35	.63	.03	36.9	43.9	88.4
Mean											
Selected	89.0	69.0	.45	.36	.40	.34	.65	.30	40.4	30.7	89.1
Non-selected	91.0	69.2	.55	.46	.33	.28	.66	.48	47.4	26.1	84.8

[a]Point biserial correlation.

classification were resolved by checking the passages for mention of the concept being tested. In comparing the results on the PASS and TEST groups the average item difficulty and the average reliability estimates were higher for the PASS group, indicating that some learning occurred from reading the passages.

For each of the eight passages/tests a core of ten K and ten CAA items had to be selected for analysis in the final study. Of these ten CAA items four were comprehension, four were application and two were analysis. A mean item difficulty of .50 and discrimination index of .30 were desired. Table 2 presents the difficulty and discrimination levels and inter-rater agreement percentages for these core items for both the PASS and TEST groups. Since the PASS condition was more similar to the final study than the TEST condition, the difficulty and discrimination indices for the PASS group were the better indicators of what might happen in the final study. For the PASS group the core K items had an average difficulty of .50 and discrimination of .42. On the CAA items the average difficulty was .36 and the discrimination was .41. In general the more difficult and more discriminating CAA items were retained as the core items for the final study. In fact, the greater overall difficulty of all CAA items made it impossible to select items with a mean difficulty of .50. Some item alternatives were revised in order to improve the difficulty and discrimination levels for the final study. The indices in Table 2 do not reflect these changes.

Final Study

Subjects

A total of 288 eleventh grade students from three high schools volunteered to participate in the study. These same schools also participated in the pilot

Table 2

Item Analysis Data for PASS and TEST Groups on the Core K and CAA Items for Each of the Selected Passages

Passage/Test	K Items					CAA Items				
	PASS		TEST		Rater	PASS		TEST		Rater
	Diff	Disc	Diff	Disc	Agree	Diff	Disc	Diff	Disc	Agree
Science										
Atmosphere	.46	.53	.36	.54	100	.39	.32	.31	.30	90
Mars & Venus	.52	.38	.31	.37	100	.37	.47	.34	.46	90
Nervous System	.49	.44	.32	.38	100	.38	.42	.33	.42	100
Sound	.48	.37	.38	.46	90	.41	.31	.34	.40	100
Mean	.49	.43	.53	.44	97.5	.39	.38	.33	.39	95.0
Social Studies										
Feed										
Population	.56	.38	.36	.33	90	.36	.38	.40	.39	100
Kinship	.51	.43	.32	.17	70	.38	.53	.39	.40	90
Money	.52	.29	.24	.35	100	.27	.52	.20	.41	90
USSR Political	.44	.56	.29	.27	90	.33	.35	.30	.45	90
Mean	.51	.42	.40	.28	87.5	.33	.44	.34	.31	92.5
All Passages/ Tests	.50	.42	.46	.36	92.5	.36	.41	.33	.35	93.7

study, although the final study was conducted the following year. The study was conducted either on a non-school day or after school, and each student was paid $10 for his participation.

Design

The design was factorial and included five factors: type of test (Memory - K items; Higher - CAA items; Combination - KCAA items), sex, ability level (IQ groups: 80-97,98-105,106-114, 115-140), subject matter sequence (science or social studies passages first), and session (a repeated measures factor with five levels). All subjects read five different passages, taking a test after each passage. Subjects in the Memory treatment had tests composed of only K items on the first three passages, but had both K and and CAA test items on the last two passages. Those in the Higher treatment had tests composed of only CAA items on the first three passages, but both K and CAA items on the last two passages. Combination subjects had both K and CAA items on all five tests. This basic part of the design is illustrated in Figure 1. The intent of the design was to establish a set for specific types of items during the first three sessions. The fourth and fifth sessions tested the ability of the subject to break the set that was presumably developed in the Memory and Higher treatments during the first three sessions by including both types of items on the fourth and fifth tests. The Combination condition served mainly as a control.

For the subject matter sequence factor half the subjects read science topics on the first four sessions and then switched to a social studies passage on the last session. The other half had the reverse subject matter sequence.

The four ability groups were based on the students' ninth grade Otis-Lennon intelligence test scores which were obtained from school records. Subjects were randomly assigned by ability level and sex to treatment and subject matter sequence combinations. Within these conditions the specific order of the passages was random for each subject. Excluding the repeated measures factor of session, there were six subjects per cell.

Experimental Groups	Type of Test	
Memory	K Items	KCAA Items
Higher	CAA Items	KCAA Items
Combination	KCAA Items	KCAA Items
	1 2 3	4 5
	Session	

Figure 1. Basic design of the study.

Tests

All tests were composed of twenty multiple-choice items. Filler items were added to the core K and core CAA items when necessary in order to have all tests of equal length. The Combination condition had the core K and CAA items on all five sessions. The Memory group had 20 K items on the first three sessions (10 core and 10 filler). The Higher group had 20 CAA items on the first three sessions (10 core and 10 filler), All groups then had the core 10K and 10 CAA items on the last two sessions. The balance among the Comprehension, Application and Analysis filler items was similar to that among the core CAA items.

Procedure

Each student read a passage and then took the corresponding test. Monitors scored each student's test as soon as he was finished, so that immediate

feedback (# correct) was provided. The student was then given the next set of materials. Students were allowed to progress at their own rate, with the total time ranging from 2½ to 4½ hours. Students were asked to record their starting and completion times for each passage. After completion of all passages and tests, each student was informed of the purpose of the study.

Results

Since passage per se was not a factor in the design, the dependent variable was a score which represented a sum over all passages. In order to eliminate unequal test variance in this sum, the raw scores for each test on the core K and CAA items were transformed to T scores. The means and standard deviations used for these transformations were based on all five sessions, so that the relative differences within factor levels would be retained. Table 3 presents these overall means and standard deviations on raw scores for the core items on each test. Results on filler items are presented for comparative purposes. Both the core K and CAA items were more difficult than the filler items, although the variability of core and filler items was similar.

Four factorial analyses of variance were conducted: two on the core K items, comparing the Memory and Combination groups on the first three sessions and all groups on the last two sessions; and two on the CAA items, comparing the Higher and Combination groups on the first three sessions and all groups on the last two sessions. Tables 4 and 5 present these analyses.

For each analysis strong ability differences occurred (Tables 4 and 5) with the performance of the ability groups on both the K and CAA items ordering the same as their group IQ scores. These means and standard deviations are given in Table 6.

Table 3

Means and Standard Deviations over the Five Sessions for the Core and Filler Items

Test	K Items		CAA Items	
	M	SD	M	SD
Science-Core				
Atmosphere	4.45	2.34	3.77	1.75
Mars & Venus	5.18	2.21	4.22	2.03
Nervous System	4.95	2.21	3.62	1.99
Sound	4.79	2.10	3.87	1.73
Mean				
Core	4.48	2.21	3.87	1.87
Filler	5.66	1.79	4.36	1.98
Social Studies-Core				
Feed Population	5.46	2.22	3.39	1.57
Kinship	5.53	2.31	4.26	2.00
Money	4.72	2.17	2.73	1.41
USSR Political	4.46	2.14	3.37	1.63
Mean				
Core	5.04	2.21	3.44	1.65
Filler	5.91	2.07	4.47	1.69
Mean (All tests)				
Core	4.94	2.21	3.66	1.76
Filler	5.78	1.93	4.41	1.83

Table 4

Analysis of Variance on the Core K Items
(A:Type of Test, B:Ability, C:Sex, D:Subject Matter Sequence, E:Session)

Source	Sessions 1-3			Sessions 4-5		
	df	MS	F	df	MS	F
Between Ss						
A	1	11.11	.09	2	237.21	2.20
B	3	4,476.02	37.13***	3	3,822.64	35.47***
C	1	17.36	.14	1	47.26	.44
D	1	24.17	.20	1	70.14	.65
AB	3	83.50	.69	6	94.32	.87
AC	1	1.00	.01	2	553.69	5.14**
AD	1	21.01	.17	2	14.64	.13
BC	3	214.94	1.78	3	87.48	.81
BD	3	43.60	.36	3	58.33	.54
CD	1	315.06	2.61	1	.92	.01
ABC	3	31.57	.26	6	53.62	.50
ABD	3	192.28	1.59	6	42.49	.39
ACD	1	8.51	.07	2	28.13	.26
BCD	3	95.85	.79	3	12.84	.12
ABCD	3	231.62	1.92	6	75.15	.70
Error	160	120.54		240	107.75	
Within Ss						
E	2	53.26	.92	1	187.92	2.83
AE	2	48.63	.84	2	21.23	.32
BE	6	36.91	.64	3	11.48	.17
CE	2	2.41	.04	1	116.46	1.75
DE	2	5.54	.09	1	7.79	.12
ABE	6	82.26	1.42	6	19.38	.29
ACE	2	29.26	.51	2	3.26	.05
ADE	2	217.83	3.75*	2	31.34	.47
BCE	6	73.89	1.27	3	30.31	.46
BDE	6	51.20	.88	3	5.90	.09
CDE	2	50.92	.88	1	279.17	4.20*
ABCE	6	38.05	.66	6	73.52	1.11
ABDE	6	44.47	.77	6	74.94	1.13
ACDE	2	55.89	.96	2	8.21	.12
BCDE	6	36.88	.63	3	91.38	1.38
ABCDE	6	67.22	1.16	6	113.89	1.71
Error	320	58.00		240	66.39	

*** $p < .001$
** $p < .01$
* $p < .05$

Table 5

Analysis of Variance on the Core CAA Items
(A:Type of Test, B:Ability, C:Sex, D:Subject Matter Sequence, E:Session)

Source	Sessions 1-3			Sessions 4-5		
	df	MS	F	df	MS	F
Between Ss						
A	1	89.46	.76	2	25.81	.24
B	3	3,400.13	28.93***	3	1,889.57	17.85***
C	1	118.26	1.01	1	48.42	.46
D	1	104.21	.89	1	2.64	.02
AB	3	119.23	1.01	6	63.20	.60
AC	1	435.76	3.71	2	205.27	1.94
AD	1	29.79	.25	2	264.41	2.50
BC	3	65.31	.55	3	64.80	.61
BD	3	71.39	.61	3	32.42	.31
CD	1	6.46	.05	1	213.89	2.02
ABC	3	157.45	1.34	6	13.55	.13
ABD	3	222.73	1.89	6	84.89	.80
ACD	1	26.26	.22	2	111.79	1.05
BCD	3	286.62	2.44	3	76.82	.72
ABCD	3	225.43	1.92	6	132.07	1.25
Error	160	117.54		240	105.84	
Within Ss						
E	2	120.10	1.73	1	100.83	1.22
AE	2	71.22	1.02	2	3.23	.04
BE	6	38.39	.55	3	12.80	.15
CE	2	19.07	.27	1	18.42	.22
DE	2	45.28	.65	1	58.14	.71
ABE	6	19.97	.29	6	53.80	.65
ACE	2	163.47	1.35	2	149.98	1.82
ADE	2	46.23	.66	1	138.01	1.68
BCE	6	18.34	.26	3	51.88	.63
BDE	6	127.28	1.83	3	44.93	.55
CDE	2	110.19	1.58	1	71.54	.87
ABCE	6	110.36	1.59	6	117.67	1.43
ABDE	6	231.12	3.32**	6	78.28	.95
ACDE	2	15.57	.22	2	39.30	.48
BCDE	6	29.96	.43	3	111	1.36
ABCDE	6	54.99	.79	6	63.00	.77
Error	320	69.58		240	82.25	

*** $p < .001$
** $p < .01$

Table 6

Means and Standard Deviations for
the Four Ability Groups on the K and CAA Items

Items		Ability Groups			
		High	Upper Middle	Lower Middle	Low
K Items (Sessions 1-3)	M	58.61	51.25	48.65	45.55
	SD	8.34	8.67	8.78	8.61
K Items (Sessions 4-5)	M	55.98	49.45	46.87	43.88
	SD	9.94	8.79	8.37	8.92
CAA Items (Sessions 1-3)	M	56.90	51.33	47.87	45.79
	SD	10.12	9.32	9.07	8.30
CAA Items (Sessions 4-5)	M	54.26	50.98	48.09	45.90
	SD	9.60	9.52	9.71	8.64

On the K items the only additional significant between subjects effect was a type of test by sex interaction on the last two sessions ($F = 5.14$, $df = 2/240$, $p < .01$). This interaction is illustrated in Figure 2. For males the Combination treatment produced the highest K scores (52.2 versus 47.9), while for females the Memory treatment was the best (50.3 versus 47.9). Performance by both sexes in the Higher treatment, by males in the Memory treatment, and by females in the Combination treatment were all at the same low level (47.9). These means are given in Table 6.

Two significant interactions occurred on the K items which included the session (within subjects)factor. On the first three sessions, a type of test by subject matter sequence by session interaction was significant ($F = 3.75$, $df = 2/320$, $p < .05$). Performance on both the Memory and Combination science sequences was relatively steady across the three sessions (maximum variation in means was 1.6 points); while the Combination social studies sequence produced steadily decreasing performance (from 53.0 to 48.9) and the Memory social studies sequence produced an initial two point increase in performance and then leveled off (Figure 3). On the last two sessions, the sex by subject matter sequence by session interaction was significant ($F = 4.2$, $df = 1/240$, $p < .05$). Since the design required a change in subject matter from the fourth to the fifth or last session, those subjects in the science sequence changed from a science passage on session 4 to a social studies passage on session 5. The opposite occurred within the social studies sequence. The interaction indicated that females within the social studies sequence decreased 3.7 points in performance from the fourth to the fifth session (Figure 4), while males within the same sequence increased only one point in performance. Both males and females within the science sequence dropped slightly from the fourth to the fifth session (1.4 and .4 points respectively).

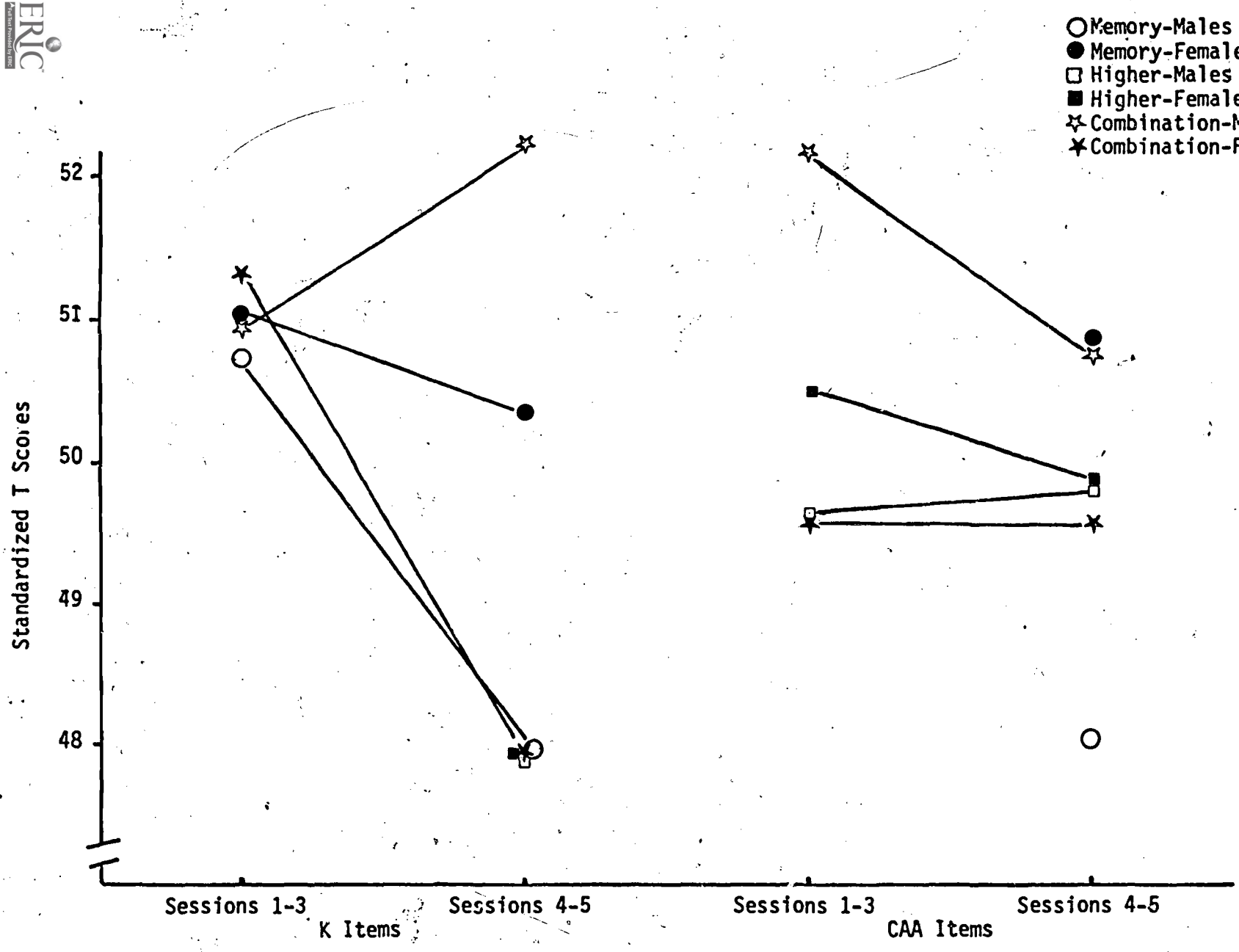


Figure 2. Means on the K and CAA items for males and females within each test condition on sessions 1-3 and 4-5.

Table 6

Means and Standard Deviations on K and CAA Items
for Sex, Type of Test, and the First and Last Sessions

Type of Test - Sex	K Items				CAA Items			
	Sessions 1-3		Sessions 4-5		Sessions 1-3		Sessions 4-5	
	M	(SD)	M	(SD)	M	(SD)	M	(SD)
Combination - Male	50.94	(10.39)	52.25	(11.24)	52.19	(10.38)	50.75	(10.74)
Combination - Female	51.37	(9.64)	47.98	(8.44)	49.54	(10.06)	49.56	(10.40)
Higher - Male	---		47.84	(10.46)	49.66	(10.14)	49.82	(10.90)
Higher - Female	---		47.95	(10.75)	50.50	(9.76)	49.89	(9.39)
Memory - Male	50.75	(10.03)	47.91	(9.66)	---		48.00	(7.68)
Memory - Female	51.01	(9.39)	50.34	(9.39)	---		50.85	(9.71)

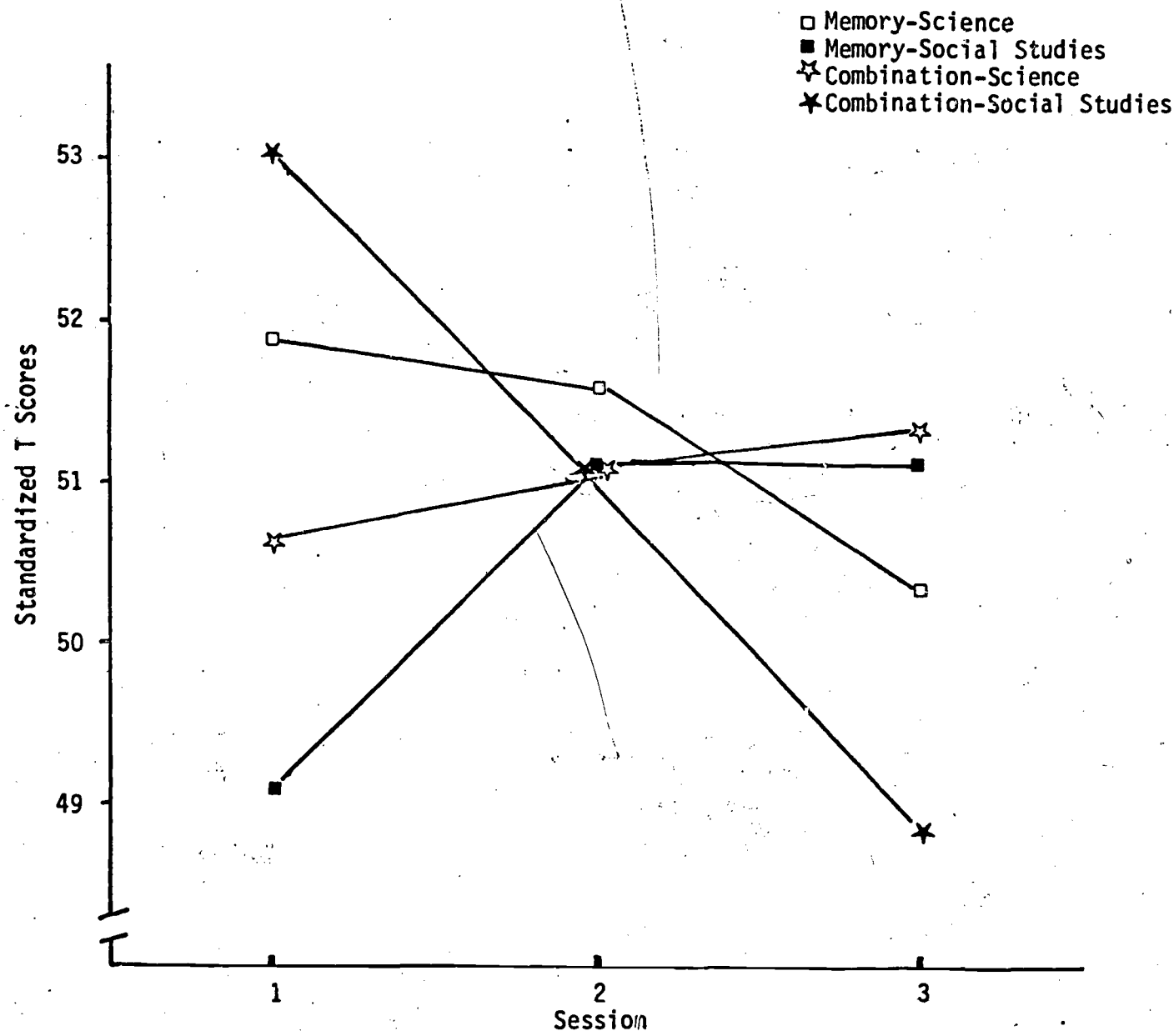


Figure 3. Interaction between type of test, subject matter sequence, and sessions 1 through 3 on the K items.

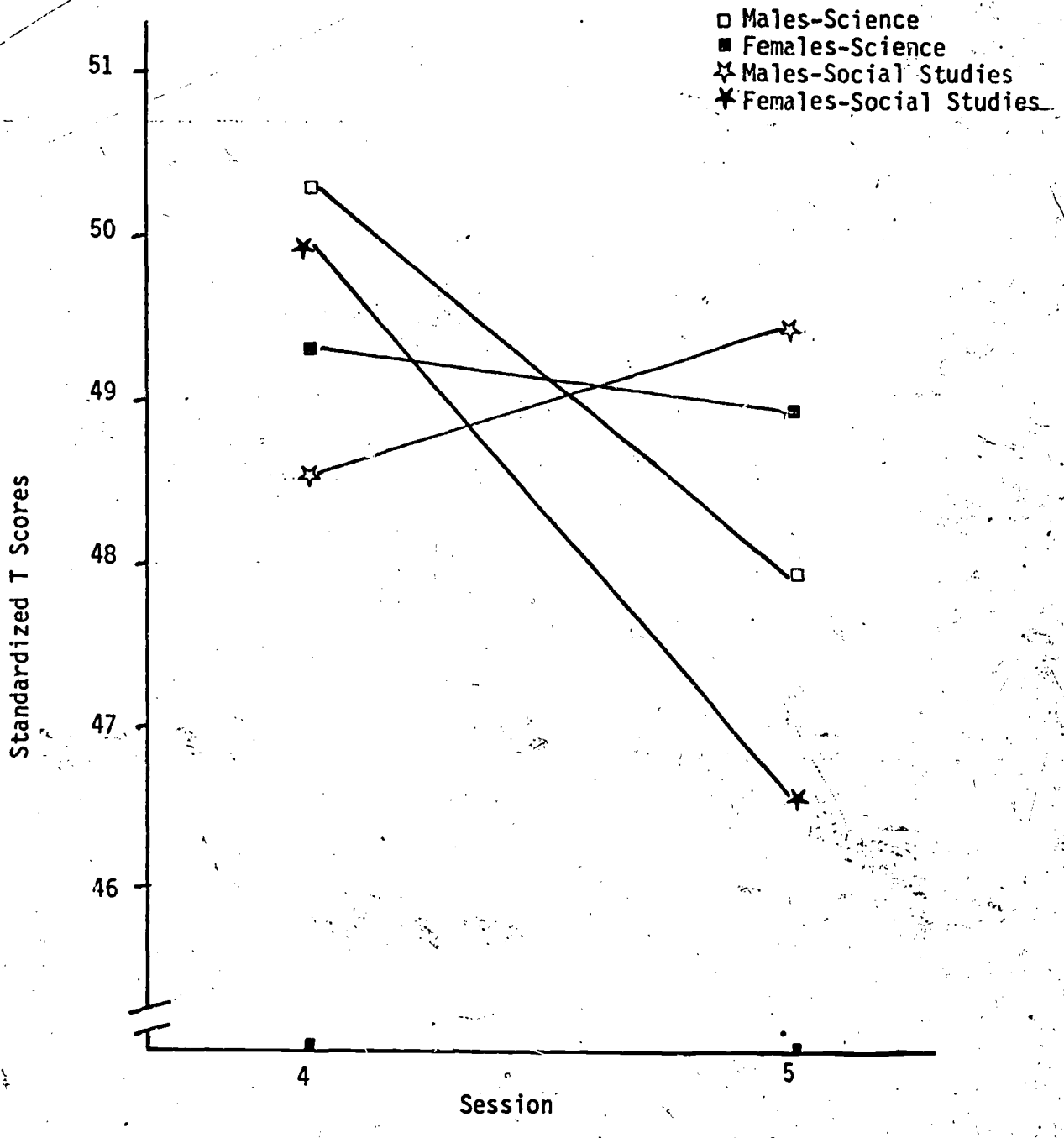


Figure 4. Interaction between sex, subject matter sequence, and sessions 4 and 5 on the K items.

Other than ability differences the only significant effect on the CAA items was a four-way interaction between type of test, ability, subject matter sequence, and the first three sessions ($F = 3.32$, $df = 6/320$, $p < .01$). The strongest discrepancies occurred for the two highest ability groups in the Combination condition. Within the science sequence the performance by the high ability group increased over sessions (from 59 to 62.7) and the upper middle group decreased (from 53.6 to 46.2). However, for the social studies sequence the reverse occurred, with the high ability group decreasing (60.7 to 51.3) and the upper middle group increasing (46.5 to 55). The performance by the other ability groups was rather steady over the three sessions. For the Higher condition, the high and two middle ability groups were relatively steady for both sequences, while the low ability group decreased within the science sequence (49 to 39.7) and increased within the social studies sequence (43.6 to 48.7).

Since the CAA items were particularly difficult for all subjects (on the average only 37% of the students answered each item correctly), the magnitude of some effects on these items could have been reduced. Thus effects significant at the .10 level and which also replicated significant effects at the .05 level for the K items were examined. Only one such effect occurred, a typeof test by sex interaction on the first three sessions ($F = 3.71$, $df = 1/160$). Again the males did best under the Combination treatment (52.2 versus 49.7); while the Higher treatment was best for females (50.5 versus 49.5), although females did not perform as high as the Combination male group (Figure 2).

Correlations among the item types were examined. The K items intercorrelated higher than the CAA items. The average correlation among the K items was .44 for Combination and .40 for Memory; while the average correlation among the

CAA items was .31 for Combination and .28 for Higher. Intercorrelations between the K and CAA items were .37 for Combination, .35 for Higher, and .29 for Memory.

The time required to read each passage was analyzed separately since it was not independent of the treatment conditions and could not be used as a covariate. Significant effects occurred for sessions ($F = 17.54$, $df = 4/860$, $p < .001$) and the type of test by ability by sex interaction ($F = 2.49$, $df = 6/240$, $p < .05$). The time required to read the passages decreased with each session (from 17.8 to 14.5 minutes). The interaction indicated more variability for males (range of cell means was from 10.9 to 19.2 minutes) than for females (15.4 to 18.5 minutes). The times for the high ability Combination males and the upper middle ability Higher males were rather fast (10.9 and 11.9 minutes respectively). Neither of these effects on time coincided with the treatment effects on K and CAA achievement.

Discussion

Contrary to expectation a simple application of the psychology of set did not adequately predict the major outcomes of the study. That is, the Memory treatment did not produce the highest performance on the K items, nor did the Higher treatment produce the highest performance on the CAA items. As expected, differences among the ability groups did occur.

Perhaps the treatment was too short and previously established sets and learning strategies were operative. The lack of steady improvement over the sessions and the strong ability effects support the notion that a facilitating set was not established by the Memory and Higher conditions for their respective item types. In addition, if the results are viewed in the larger context of a set created by memory type tests given in the school, then presumably all students entered the experiment with a set to study for memory type items, and such a set

could have inhibited their ability to perform well on the higher-level items. The high difficulty level of the CAA items in both the pilot and final study provides support for this hypothesis.

The interactions between sex and type of test were generally in the same direction on both K and CAA items. Males in the Combination condition scored highest on both K and CAA items, while those in the Memory condition scored the lowest. The only exception to this ordering was the similarity of males in both Memory and Combination groups on the K items in the first three sessions. On the other hand, only the Memory treatment really facilitated female performance on both the K and CAA items. The Higher condition did not facilitate performance by either sex on the K items, although some facilitation did occur on the CAA items. The ineffectiveness of the Higher treatment for both sexes can probably be attributed to the concentrated exposure to the rather difficult items provided under this condition. Difficult items have been shown (Marso, 1969) to inhibit further learning.

Why was female performance facilitated by memory items and male performance by a variety of item types? Studies have indicated that females prefer mnemonic to logical or choice learning strategies (Gay, 1972; Goldman, 1972), and that females perform better on direct recall of written material (Todd & Kessler, 1971). If it is assumed that most of the students had a set to study for memory type items, and therefore that the Memory condition was the one most similar to the school situation, then the greater inability of females (as opposed to males) to break a set could have contributed in part to the higher performance of females on the K items. In addition, it is likely that the Combination condition represented a challenging, but not totally discouraging, situation. The literature on sex differences has shown that males generally possess a more autonomous

approach to school than females (Coleman, 1961) as indicated by the fact that males "are likely to do well in subjects that interest them and poorly in subjects that bore them (Maccoby, 1966, p.32)," and that "boys are more likely to rise to an intellectual challenge, girls to retreat from one (Maccoby, 1966, p.33)." These factors may partly explain why males performed well under the Combination condition.

It is difficult to generalize from this experimental setting to a classroom setting where a much longer exposure to tests could be provided, where grades affect the motivation level of students, and where students would be more likely to thoroughly study the subject matter. In addition, the other significant interactions which included the session and subject matter sequence factors make generalization to other situations dependent upon the length of the experiment and the content used. However, the data indicate that future studies of either a basic or more applied nature shouldinclude such factors in order to clarify the effects of tests upon cognitive processes.

References

Billings, G. Cognitive levels of elementary science tests. School Science and Mathematics, 1971, 71, 824-830.

Bloom, B. S. (Ed.) Taxonomy of educational objectives: Cognitive domain. New York: David McKay, 1956.

Coleman, J. S. The adolescent society: The social life of the teenager and its impact on education. New York: Free Press of Glencoe, 1961.

Cooper, J. M. Two types of social studies examinations and their effect on student learning. Unpublished doctoral dissertation, Stanford University, 1967. (University Microfilms #67-17,404)

Davis, O. L., & Hunkins, F. P. Textbook questions: What thinking processes do they foster? Peabody Journal of Education, 1966, 43, 285-292.

Gall, M. D. Use of questions in teaching. Review of Educational Research, 1970, 40, 707-721.

Gay, L. R. Use of a retention index for mathematics instruction. Journal of Educational Psychology, 1972, 63, 466-472.

Goldman, R. D. Effects of a logical versus a mnemonic learning strategy on performance in two undergraduate psychology classes. Journal of Educational Psychology, 1972, 63, 347-352.

Hunkins, F. P. The effects of analysis and evaluation questions on various levels of achievement. Paper presented at the annual meeting of the American Educational Research Association, Chicago, February 1968.

Jeffrey, J. M. Psychological set in relation to the construction of mathematics tests. The Mathematics Teacher, 1969, 62, 636-638.

Johnson, D. M. The psychology of thought and judgment. New York: Harper, 1955.

Luchins, A. S., & Luchins, E. H. Rigidity of behavior: A variational approach to the effect of einstellung. Eugene, Oregon: University of Oregon Books, 1959.

McKenzie, G. R. Some effects of frequent quizzes on inferential thinking. American Educational Research Journal, 1972, 9, 231-240.

Maccoby, E. E. (Ed.) The development of sex differences. Stanford, California: Stanford University Press, 1966.

Marso, R. N. The influence of test difficulty upon study efforts and achievement. American Educational Research Journal, 1969, 6, 621-632.

Ray, W. S. The experimental psychology of original thinking. New York: Macmillan, 1967.

Rothkopf, E. Z. The concept of mathemagenic activities. Review of Educational Research, 1970, 40, 325-336.

Todd, W. B., & Kessler, C. C. Influence of response mode, sex, reading ability, and level of difficulty on four measures of recall of meaningful written material. Journal of Educational Psychology, 1971, 62, 229-234.

Watts, G. H., & Anderson, R. C. Effects of three types of inserted questions on learning from prose. Journal of Educational Psychology, 1971, 62, 387-394.