R E P O R T   R E S U M E S

ED 012 803 24 AA 000 163

REINFORCEMENT IN CLASSROOM LEARNING. PART II, STUDIES OF REINFORCEMENT IN SIMULATED CLASSROOM SITUATIONS. PART III, IDENTIFICATION OF REINFORCERS OF HUMAN BEHAVIOR.

BY- TRAVERS, ROBERT M.W. AND OTHERS

UTAH UNIV., SALT LAKE CITY, BUREAU OF EDUC. RES.

REPORT NUMBER CRP-1196-PTS-2-3 PUB DATE SEP 64
CONTRACT OEC-2-10-010
EDRS PRICE MF-$1.50 HC-$15.80 395P.

DESCRIPTORS- *REINFORCEMENT, *LEARNING PROCESSES, *MODELS, *STIMULUS BEHAVIOR, *LEARNING MOTIVATION, OBSERVATION, INTERACTION, FEEDBACK, ATTENTION, ROTE LEARNING, PAIRED ASSOCIATE LEARNING, CREATIVE DEVELOPMENT, REWARDS, STIMULUS DEVICES,

REINFORCEMENT CONCEPTS DERIVED LARGELY FROM RESEARCH OF SUBHUMAN SUBJECTS WERE TESTED FOR APPLICABILITY TO HUMAN-LEARNING SITUATIONS SIMILAR TO THOSE THAT OCCUR IN SCHOOLS. A SERIES OF EXPLORATORY STUDIES CONDUCTED IS DESCRIBED IN PART II OF THIS REPORT. IN PART III, TWO EXPERIMENTS CONDUCTED TO DETERMINE THE REINFORCING VALUE OF DIFFERENT STIMULI ARE REPORTED. ELEMENTARY SCHOOL CHILDREN WERE THE SUBJECTS OF ALL OF THESE STUDIES AND EXPERIMENTS. THE EXPERIMENTS REPORTED IN PART II DEALT WITH THE FOLLOWING SUBJECTS--(1) A COMPARISON OF LEARNING UNDER DIRECT REINFORCEMENT WITH LEARNING UNDER VICARIOUS REINFORCEMENT, (2) LEARNING AS A CONSEQUENCE OF THE LEARNER'S TASK INVOLVEMENT UNDER DIFFERENT CONDITIONS OF FEEDBACK, (3) THE RELATIONSHIP OF LEARNING TO ATTENTION IN A SIMULATED CLASSROOM SITUATION, (4) THE EFFECTIVENESS OF PUPILS AS REINFORCING AGENTS, (5) THE EFFECT OF PUPIL-PUPIL REINFORCEMENT IN MISMATCHED PAIRS, AND (6) THE ABILITY OF PUPILS TO USE POSITIVE AND NEGATIVE INFORMATION DERIVED FROM OBSERVING THE BEHAVIOR OF OTHER PUPILS. THE TWO EXPERIMENTS REPORTED IN PART III WERE (1) THE EFFECTS OF DIFFERENT REINFORCERS, A COMPARISON ACROSS AGE LEVELS, AND (2) A COMPARISON OF THE REINFORCING EFFECT OF DIFFERENT STIMULI IN A PAIRED-ASSOCIATE LEARNING TASK. RESULTS OF ONE EXPERIMENT SHOWED THAT ALTHOUGH AN ELECTRIC SHOCK REINFORCEMENT FOR INCORRECT RESPONSES CAUSED RATS TO LEARN SIGNIFICANTLY FASTER, CHILDREN DID NOT. FROM THE STUDIES, THE AUTHOR CONCLUDED THAT (1) CHILDREN APPEAR TO BE MUCH MORE CAPABLE OF LEARNING FROM THEIR ERRORS THAN SOME PSYCHOLOGISTS HAVE SUPPOSED THEM TO BE, AND (2) REINFORCING EVENTS SHOULD SUPPLY INFORMATION IN THE CLEAREST, MOST UNAMBIGUOUS FORM. A RELATED REPORT IS ED 003 055. (AL)

1196

Reinforcement in Classroom Learning

ROBERT M. W. TRAVERS, Principal Investigator

PART II: STUDIES OF REINFORCEMENT IN SIMULATED CLASSROOM SITUATIONS

Contributors

DANIELLE H. HAYGOOD, MARY L. McCORMICK, KENT E. MYERS, IAN E. REID
EVE M. SANFORD, KEITH R. VAN WAGENEN, FRANK E. WILLIAMS

PART III: IDENTIFICATION OF REINFORCERS OF HUMAN BEHAVIOR

Contributors

MADGE N. ALTER, FRANK B. NELSON, IAN E. REID

Final Report, Parts II and III — September 1964

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
OFFICE OF EDUCATION CONTRACT NO. 2-10-010

University of Utah
Bureau of Educational Research

REINFORCEMENT IN CLASSROOM LEARNING

PART II: STUDIES OF REINFORCEMENT IN SIMULATED CLASSROOM SITUATIONS

Principal Investigator

Robert M. W. Travers

Contributors

Danielle H. Haygood
Mary L. McCormick
Kent E. Myers
Ian E. Reid
Eve M. Sanford
Keith R. Van Wagenen
Frank E. Williams

PART III: IDENTIFICATION OF REINFORCERS OF HUMAN BEHAVIOR

Principal Investigator

Robert M. W. Travers

Contributors

Madge N. Alter
Frank B. Nelson
Ian E. Reid

Final Report, Parts II and III September 1964

U. S. Department of Health, Education and Welfare

Office of Education Contract No. 2-10-010

University of Utah

Bureau of Educational Research

pupils, teachers and principals is gratefully acknowledged. They are the troops on the front line of educational research and advances are highly dependent upon their cooperation. Appreciation must also be expressed to Superintendant Harold Holt and to Dr. Arthur E. Arnesen, both of whom graciously gave of their time in order that the research workers might have access to schools within their respective districts. Appreciation is also expressed to Virginia Knowlton for editorial assistance on the final report.

Since Part I of this report provided an extensive list of relevant references, relatively few studies are cited and referenced in the chapters that follow.

TABLE OF CONTENTS

LIST OF FIGURES

SECTION II

FIGURES PAGE

SECTION III

FIGURES PAGE

LIST OF TABLES

SECTION III

TABLE PAGE

SECTION III

CHAPTER I

THE PROBLEM AND THE RESEARCH TECHNIQUE FOR STUDIES REPORTED IN CHAPTERS II THROUGH VII

The review of research related to reinforcement presented in the Section I of this report is based largely on studies of subhuman subjects, but enthusiasm for applying the knowledge thus derived to problems of education has not been accompanied by experimentation to demonstrate that such knowledge is applicable. The studies which follow in this section of the report are explorations of the applicability of reinforcement concepts to the planning of human learning situations similar to those that occur in schools.

The general purposes of these studies were the following:

1. To relate the degree of acquisition of the task to reinforcing conditions in situations similar to those occurring in the classroom.
2. To determine the amount of learning which takes place when one pupil observes another pupil in a recitation type situation.
3. To determine the effectiveness of various reinforcing statements on the part of the teacher in a recitation type of situation.
4. To compare reinforcements provided by written materials with the pupil working by himself with reinforcements provided in recitation-type situations.
5. To compare the effectiveness of learning situations in which the teacher provides the reinforcement in comparison with situations in which pupils reinforce the work of pupils.
6. To compare the effects of pupils of different achievement levels functioning in teacher roles in relation to other pupils.
7. To relate the effectiveness of the teacher's reinforcements to the distance of the teacher from the pupil.

These are the broad questions for which answers were sought, but more specific questions will be asked at the beginning of each section of the report that follows.

Method

The Task

The first step which had to be undertaken in the development of this study was the construction of a learning task which (a) could be administered in a learning situation conforming to the response-reinforcement learning model, (b) be similar to learning tasks commonly observed in classrooms, (c) be such that it could be administered under several different learning and reinforcement conditions, and (d) permit a reliable measure of learning. The task selected was that of learning the meaning of 60 German words.

The learning task consisted of a German stimulus word which in the experiment was to be correctly associated with one of two English noun responses. German was selected because it was assumed that there would be few children with any familiarity with this language, and in addition, that interest in learning foreign words would be high.

The general format of the task required a subject to guess which one of two English words meant the same as a given German word. Reinforcement followed the response. These pairs of English nouns (one of which was the equivalent of the German stimulus) had to be selected so that one could assume that they had equal association to the German stimulus. In providing a pool of such words, all nouns of three to eight letters in length (not having some obvious association with the

German word) were taken from the Thorndike-Lorge (1944, appendix Part V) list of "Original Thorndike Words." The pairs were selected separately for the first 500 most frequently occurring and for the second 500 most frequently occurring words. An additional seven were taken from the list of 110 appended to the original 1000, which yielded a total of 314 words from which the task could be assembled.

On the basis of an argument advanced by Underwood and Schulz (1960), a second step in providing for equal association value was the assignment of a numerical frequency value for the first bigram of each selected word, i.e., (ap)ple. In the pairing of response choices the bigrams of the two words had to have approximately comparable frequencies (as determined by the compilers, Underwood and Schulz).

In addition to the procedures already outlined, each English noun was checked for level of readability to insure that words used were well within the reading capacity of participating Ss. Durrell, in his Appendix A of Improving Reading Instruction (1956), lists 754 words compiled at the primary grade level as vocabulary for older children in remedial reading programs. Each word has been assigned a value ranging from 1 to 7 which indicates frequency of occurrence in the literature of small children. It was found that 69 per cent of the response nouns in the learning task are listed among Durrell's words. As far as was possible, pairs of response choices were matched with respect to Durrell's readability values. Those nouns already selected for the task which were not part of the Durrell list were checked against another Durrell list of vocabulary to be learned in the fourth grade. While not many of the learning task nouns were found there, it was obvious that Durrell's

fourth grade list was at a higher level of difficulty. It was concluded that the words selected would be well within the range of reading ability for the proposed sample of subjects.

The task was prepared in two different forms. In one form, suitable for administration in a group situation, each German word and the accompanying English words were printed in large block letters on sheets of card 11 x 14 inches. In another form, the materials were typed on paper suitable for insertion into a teaching machine (Koncept-o-graph No. 7).

The acquisition trials were conducted on each of three days: Monday, Tuesday and Wednesday. On each of these days a set of twenty words was presented with four trials on each. The total of eighty presentations per day required approximately forty minutes to complete. The criterion measure was not administered until Friday. The test called for recognition of the English equivalent among four choices. The three distractors included the words used as an incorrect choice in the learning task and also two English words derived from other task items.

The German words and the characteristics of the English equivalent and the alternative response word are given in Table 1.01. The order in which the words were given and the positions of the correct and incorrect alternatives are given in Table 1.02.

Measurement of Acquisition

Throughout the series of studies which follow, the three main sections of the learning task were administered on Monday, Tuesday and Wednesday, and the pupils were tested for retention Friday of the same week. Originally the plan had been to use a recall test in which the

Table 1.01

LEARNING TASK--SHOWING THE FREQUENCY VALUES
USED AS THE BASIS FOR EQUATING
RESPONSE CHOICES

TASK A--MONDAY

Stimulus		Responses						
1.	TEIL	(4) 1098 part (4) 1129 home	8.	ZUG	(4) 1028 train (4) 913 table	15.	GIPFEL	(5) 1213 top (2) 1210 man
2.	BAUM	(3) 1028 tree (-) 1117 corn	9.	VORRAT	(6) 3473 store (-) 3473 storm	16.	UFER	(7) 413 bank (3) 404 book
3.	RAD	*(-) 700 wheel (5) 852 money	10.	BEIN	*(6) 2794 leg (5) 2516 ear	17.	BERG	(5) 1003 hill (5) 1057 side
4.	HUETE	(4) 1213 today (4) 1271 light	11.	MUTZE	(5) 1055 cap (3) 1055 car	18.	STRASSE	*(5) 1721 road (7) 1689 news
5.	SPIESE	(7) 710 food (3) 761 bear	12.	RAND	(-) 1271 lip (6) 1396 hat	19.	GESICHT	(5) 420 fact (4) 413 ball
6.	VOGEL	(3) 234 bird (-) 205 king	13.	GLOCKE	(5) 761 bell (-) 721 fish	20.	FELSEN	(7) 1721 rock (5) 1488 rain
7.	LEIB	(6) 404 body (3) 397 door	14.	KNABE	(2) 404 boy (5) 598 sun			

Table 1.01 (Continued)

TASK B--TUESDAY

Stimulus		Responses			
1.	KUNST	*(-) art	2389	(5) arm	2389
2.	MEER	*(6) sea	1865	(-) row	1721
3.	HIMMEL	*(-) sky	111	(4) egg	295
4.	TASSE	(7) cup	459	(-) age	588
5.	LUFTEN	(6) air	829	(7) son	920
6.	GESETZ	*(-) law	1307	(-) ice	1309
7.	KRIEG	*(6) war	723	(-) bee	761
8.	EICHE	(-) oak	66	(-) oil	113
9.	TASCHE	(5) bag	413	(7) fat	420
10.	TAG	*(2) day	342	(-) bow	404
11.	KOPF	(3) head	2897	(6) east	2516
12.	ZEIT	*(2) time	2497	(-) heat	2897
13.	HOF	(5) yard	96	(-) knee	88
14.	FOLGE	(6) suit	598	(7) sail	566
15.	ZOLL	(-) inch	3902	(7) tire	2497
16.	SPIEL	(5) game	398	(7) idea	495
17.	KNOCHEN	(-) bone	404	(4) boat	404
18.	STADT	(4) town	846	(6) mile	720
19.	SCHWANZ	(-) tail	913	(7) iron	1061
20.	STUNDE	(7) hour	1129	(7) life	1271

Table 1.01 (Continued)

TASK C--WEDNESDAY

Stimulus		Response						
1.	ZIMMER	(4) 1721 room (-) 1396 hall	8.	LIED	(5) 920 song (6) 1203 ship	15.	WOLKE	(-) 648 cloud (-) 578 ocean
2.	ZEILE	(4) 1271 line (3) 1396 hand	9.	STIMME	(-) 122 voice (5) 148 floor	16.	RAUCH	(7) 90 smoke (6) 148 music
3.	HOLZ	(5) 288 wood (5) 338 wind	10.	TIER	(-) 761 beast (-) 700 wheat	17.	RAUM	*(-) 707 space (2) 723 water
4.	BRUNNEN	(2) 779 well (4) 779 week	11.	HOCHST	(-) 1297 chief (3) 1297 chair	18.	WAHRHEIT	(-) 1028 truth (-) 1129 honor
5.	AUSFLUG	(7) 1028 trip (-) 1054 lord	12.	TUCH	(7) 648 cloth (-) 648 clock	19.	MACHT	(-) 818 power (7) 703 apple
6.	HEER	(-) 2389 army (7) 3473 star	13.	ZAUN	*(-) 712 fence (5) 723 watch	20.	SACHE	(3) 2879 thing (6) 2897 heart
7.	LOCH	(5) 1129 hole (-) 1117 coal	14.	FARBE	(7) 1098 paint (-) 1117 court			

Note:

1. Response words are from "Original Thorndike Words" 1-1000.

* Items where the response did not originate in the same group of 500.

2. Values in parenthesis () 1-7 are from Durrell primary-reading-level frequencies. Number 1 indicates highest frequency.
3. Large number values are initial bigram frequencies (Underwood & Schulz).

Table 1.02

ORDER IN WHICH ITEMS WERE PRESENTED TO SUBJECTS*

TASK A--MONDAY

Order of Items on 1st and 3rd Presentation	Position of Correct R on Card	S	R	Order of Items on 2nd and 4th Presentation	Position of Correct R on Card	Subject Seating Position
1	bottom	Teil	part	30	top	4
2	top	Baum	tree	29	bottom	2
3	bottom	Rad	wheel	28	top	1
4	bottom	Heute	today	27	top	3
5	top	Speise	food	26	bottom	2
6	top	Vogel	bird	25	bottom	4
7	bottom	Leib	body	24	top	3
8	top	Zug	train	23	bottom	1
9	bottom	Vorrat	store	22	top	1
10	bottom	Bein	leg	21	top	2
11	top	Mutze	cap	40	bottom	3
12	top	Rand	lip	39	bottom	4
13	bottom	Glocke	bell	38	top	4
14	top	Knabe	boy	37	bottom	2
15	bottom	Gipfel	top	36	top	1
16	top	Ufer	bank	35	bottom	3
17	top	Berg	hill	34	bottom	3
18	bottom	Strasse	road	33	top	2
19	top	Gesicht	face	32	bottom	1
20	bottom	Felsen	rock	31	top	4

*See explanatory note bottom of next page.

Table 1.02 (Continued)

TASK B--TUESDAY

Order of Items on 1st and 3rd Presentation	Position of Correct R on Card	S	R	Order of Items on 2nd and 4th Presentation	Position of Correct R on Card	Subject Seating Position
1	top	Kunst	art	30	bottom	3
2	bottom	Meer	sea	29	top	4
3	bottom	Himmel	sky	28	top	4
4	top	Tasse	cup	27	bottom	2
5	bottom	Luften	air	26	top	1
6	top	Gesetz	law	25	bottom	3
7	top	Krieg	war	24	bottom	3
8	bottom	Eiche	oak	23	top	2
9	top	Tasche	bag	22	bottom	1
10	bottom	Tag	day	21	top	4
11	bottom	Kopf	head	40	top	4
12	top	Zeit	time	39	bottom	2
13	bottom	Hof	yard	38	top	1
14	bottom	Folge	suit	37	top	3
15	top	Zoll	inch	36	bottom	2
16	top	Spiel	game	35	bottom	4
17	bottom	Knochen	bone	34	top	3
18	top	Stadt	town	33	bottom	1
19	bottom	Schwanz	tail	32	top	1
20	top	Stunde	hour	31	bottom	2

*This list of twenty items was repeated four times (80 item presentations). Note that item 10, Bein, also becomes item 21, and is presented to the subject at position 2. Forty presentations are shown here. The same order is repeated once so that there are eighty presentations in all.

Table 1.02 (Continued)

TASK C--WEDNESDAY

Order of Items on 1st and 3rd Presentation	Position of Correct R on Card	S	R	Order of Items on 2nd and 4th Presentation	Position of Correct R on Card	Subject Seating Position
1	bottom	Zimmer	room	30	top	4
2	top	Zeile	line	29	bottom	2
3	bottom	Holz	wood	28	top	1
4	bottom	Brunnen	well	27	top	3
5	top	Ausflug	trip	26	bottom	2
6	top	Heer	army	25	bottom	4
7	bottom	Loch	hole	24	top	3
8	top	Lied	song	23	bottom	1
9	bottom	Stimme	voice	22	top	1
10	bottom	Tier	beast	21	top	2
11	top	Hochst	chief	40	bottom	3
12	top	Tuch	cloth	39	bottom	4
13	bottom	Zaun	fence	38	top	4
14	top	Farbe	paint	37	bottom	2
15	bottom	Wolke	cloud	36	top	1
16	top	Rauch	smoke	35	bottom	3
17	top	Raum	space	34	bottom	3
18	bottom	Wahrheit	truth	33	top	2
19	top	Macht	power	32	bottom	1
20	bottom	Sache	thing	31	top	4

pupils would be given the German words and be required to enter the English equivalent. However, such a test was found to be much too difficult in relation to the time devoted to learning, so a recognition test was developed which gave each German word followed by four English words. The pupil was required to choose the correct English alternative from the four words given. The four words included the pair that had been originally presented with the German word and two additional words were added from other items in the learning task. The reliability for the score derived from the test was estimated by means of the Kuder-Richardson Formula 21 and was found to be 0.72 (N=180). This estimate was derived from the cases included in the first study to be reported.

A copy of the recognition test is shown in Table 1.03. The three parts of the test were administered in the same order one after the other. The test was administered by a member of the staff of the project who read aloud to the pupils the German word and the four English words for each item and then allowed time for the pupils to choose an answer before moving on to the next item. Pupils were urged to attempt each item and very few of the pupils skipped any items. The test was scored for the number of correct responses.

Table 1.03

RECOGNITION TEST A (MONDAY'S TASK) ADMINISTERED FRIDAY

___ KNABE	___ RAND	___ SPEISE	___ GLOCKE
1. KING	1. HAT	1. FOOD	1. LEG
2. BOY	2. NIGHT	2. BED	2. ADD
3. DOOR	3. TABLE	3. FISH	3. BELL
4. SUN	4. LIP	4. BEAR	4. FISH
___ GIPFEL	___ BERG	___ VOGEL	___ ZUG
1. MAN	1. PEN	1. EYE	1. TABLE
2. SIDE	2. RAIN	2. JOB	2. HAT
3. TOP	3. HILL	3. KING	3. TRAIN
4. LIGHT	4. SIDE	4. BIRD	4. TREE
___ MUTZE	___ UFER	___ LEIB	___ HEUTE
1. HOUSE	1. BANK	1. WIFE	1. LIGHT
2. CAP	2. BOX	2. BODY	2. FOOD
3. NEWS	3. BOY	3. DOOR	3. TODAY
4. CAR	4. BOOK	4. DATE	4. NIGHT
___ BEIN	___ STRASSE	___ BAUM	___ VORRAT
1. LEG	1. COW	1. MAN	1. END
2. EAR	2. NEWS	2. HOME	2. STORM
3. BED	3. CORN	3. CORN	3. STATE
4. TIE	4. ROAD	4. TREE	4. STORE
___ GESICHT	___ FELSEN	___ TEIL	___ RAD
1. BALL	1. RAIN	1. PART	1. WHEEL
2. FACE	2. TODAY	2. HILL	2. BODY
3. PLACE	3. ROCK	3. WHEEL	3. MONEY
4. GIFT	4. CAP	4. HOME	4. PIECE

Table 1.03 (Continued)

RECOGNITION TEST B (TUESDAY'S TASK) ADMINISTERED FRIDAY

___ LUFTEN	___ KUNST	___ KOPF	___ SPIEL
1. AIR	1. ARM	1. ART	1. DAY
2. WAR	2. ART	2. HEAD	2. BOW
3. SON	3. ADD	3. TIME	3. GAME
4. BABY	4. AIR	4. EAST	4. IDEA
___ TAG	___ TASCHE	___ ZEIT	___ ZOLL
1. KNEE	1. FAT	1. TIME	1. TIRE
2. BOW	2. IDEA	2. LEG	2. STATE
3. GAME	3. BAG	3. HEAT	3. STORY
4. DAY	4. BOAT	4. TIE	4. INCH
___ MEER	___ TASSE	___ STUNDE	___ FOLGE
1. ROAD	1. CUP	1. IRON	1. BOAT
2. SEA	2. AGE	2. LIFE	2. BONE
3. NEWS	3. SKY	3. TAIL	3. SAIL
4. ROW	4. SAIL	4. HOUR	4. SUIT
___ KRIEG	___ EICHE	___ STADT	___ SCHWANZ
1. TOWN	1. OAK	1. WHEEL	1. HOUR
2. BELL	2. EGG	2. TABLE	2. TAIL
3. BEE	3. OIL	3. TOWN	3. IRON
4. WAR	4. YARD	4. MILE	4. LIFE
___ GESETZ	___ HIMMEL	___ KNOCHEN	___ HOF
1. KNEE	1. OAK	1. BOAT	1. YARD
2. IRON	2. OIL	2. BONE	2. KNEE
3. LAW	3. EGG	3. BOX	3. SKY
4. ICE	4. SKY	4. FARM	4. EGG

Table 1.03 (Continued)

RECOGNITION TEST C (WEDNESDAY'S TASK) ADMINISTERED FRIDAY

___ HOLZ	___ LOCH	___ MACHT	___ ZAUN
1. SMOKE	1. HOLE	1. APPLE	1. SPACE
2. MUSIC	2. COAL	2. POWER	2. FENCE
3. WOOD	3. PAINT	3. BEAST	3. WATCH
4. WIND	4. COURT	4. WHEAT	4. WATER
___ BRUNNEN	___ LIED	___ TIER	___ FARBE
1. WELL	1. CLOCK	1. BEAST	1. COAL
2. WEEK	2. SHIP	2. WHEAT	2. PAINT
3. CLOTH	3. CLOUD	3. OCEAN	3. COURT
4. FENCE	4. SONG	4. CLOUD	4. HOLE
___ ZEILE	___ AUSFLUG	___ HOCHST	___ RAUCH
1. HAND	1. TRIP	1. HONOR	1. SMOKE
2. RAIN	2. LORD	2. CHAIR	2. FLOOR
3. LINE	3. CHIEF	3. TRUTH	3. VOICE
4. PEN	4. TRUTH	4. CHIEF	4. MUSIC
___ ZIMMER	___ STIMME	___ TUCH	___ RAUM
1. CHAIR	1. SMOKE	1. CLOCK	1. WELL
2. HALL	2. VOICE	2. SHIP	2. WEEK
3. TRIP	3. MUSIC	3. CLOTH	3. WATER
4. ROOM	4. FLOOR	4. SONG	4. SPACE
___ HEER	___ SACHE	___ WOLKE	___ WAHRHEIT
1. STAR	1. STAR	1. OCEAN	1. LINE
2. ARMY	2. HEART	2. MONEY	2. HONOR
3. HEART	3. THING	3. WHEEL	3. TRUTH
4. THING	4. ARMY	4. CLOUD	4. HAND

CHAPTER II

A COMPARISON OF LEARNING UNDER DIRECT REINFORCEMENT WITH VICARIOUS REINFORCEMENT

In Part I of this report a review was made of studies of learning in which one organism observes another organism learn. The learning which occurs under this condition has been named vicarious learning in order to distinguish it from learning occurring under more typical laboratory conditions in which the learner responds and receives some kind of direct reinforcement or feedback. In the studies that follow the two learning conditions are referred to as the vicarious condition and the direct condition. The learning that results is referred to as vicarious learning and direct learning.

In the review of the literature, evidence was presented which substantiates the position that vicarious learning occurs in primates but there is some doubt whether it is a phenomenon in simpler organisms. Generally it appears to provide a slower form of learning than direct learning. The central problem to be investigated in this study is the extent to which vicarious learning in young human learners is comparable to learning by a direct reinforcement method. The problem is an important one in that much of the work of a typical class in an American elementary school is undertaken by a recitation method in which some pupils learn by direct reinforcement while others learn by a vicarious process. In the study that follows, learning is investigated under conditions that simulate closely those found in a recitation situation in a typical classroom. In addition, the learning data derived from such a situation will be compared with data derived from situations in which the pupil is either learning alone or is learning while interacting with a single adult. The experiment permits the comparison of the effects of a number of different reinforcing conditions.

The Reinforcing Conditions

The study included four main reinforcing conditions, two involving a group of pupils and two involving only a single pupil. First let us consider the two conditions involving pupils working in groups which will be referred to as Conditions 1 and 2.

The physical characteristics of the experimental arrangement were intended to be suggestive of a classroom situation. A "teacher"-experimenter faced eight "pupil"-subjects who were seated in a somewhat familiar classroom row (see Figure 2.01).

1. Direct Reinforcement Condition.--Four of the group Ss were separately and directly interacted with by the experimenter. Each of the four Ss was presented an equal proportion of items from the total task. One subject at a time (designated D, and numbered 1, 2, 3, or 4-- Figure 2.01) was called upon by name, asked to respond to a stimulus word, and given feedback as to the correctness of the response. For example, the experimenter might say, "John, this word is Baum" (the German stimulus word). The subject, John, would then select one of the two English alternatives as the equivalent of Baum. If the S said "tree," the correct choice, the experimenter would say, "right" or "that's right." When he was wrong the experimenter said nothing. The remainder of the group listened to the exchange. The four actively involved Ss will hereafter be referred to as direct reinforcement subjects.

There were four presentations of each card which contained a stimulus and two response choices. All four presentations were always given to the same subject. The nature of the task was such that only one S at a time could be directly involved in responding and receiving feedback. This

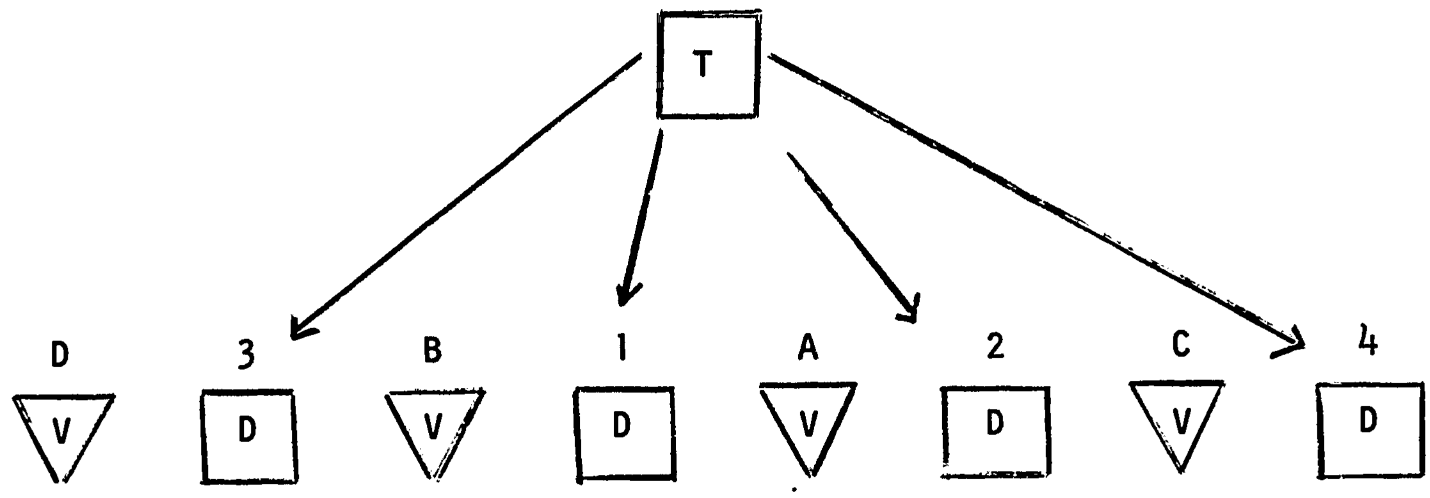


Fig. 2.01 Arrangement of teacher and pupils in the experimental recitation situation. The pupils marked D are those that interacted directly with the teacher. The V pupils did not interact with the teacher and learned vicariously.

meant that three of the four Ss would be indirectly, or vicariously involved in learning an item which was administered to the fourth.

The learning condition for the D subjects can be divided into two conditions (a) learning on those items to which they responded directly (1/4 of items), and (b) learning on those items to which the subjects did not respond directly (3/4 of the items). The same Ss serve for both treatments. Each S responds directly to 1/4 of the items, and listens to the other three Ss respond to the remaining 3/4.

2. Vicarious Reinforcement Condition.--To complete the group of eight, four additional Ss were seated alternately to those mentioned above. These are the subjects marked with V's in Figure 2.01. These were not at any time directly interacted with. Whatever learning was to occur for them, had to occur as a result of observing; that is, hearing the response of a direct S and the consequent feedback of the experimenter. This vicarious treatment (all of the items learned vicariously) differs from 3/4 vicarious above in that the V subjects are never called upon and thus may develop an expectancy of not being called upon at all.

3. Isolation Condition-Working Alone. A group of four subjects was used to represent a "working alone" condition. Subjects were seated before a "teaching machine" (Koncept-O-Graph, Model 1 7) from which visual feedback was afforded. Responses to the same task described earlier were made internally, not vocally or written. The subject simply advanced the machine by hand to the problem, subvocally responded, advanced the machine to the answer, and repeated the process for the entire task. The type of material which composed the task was not of the sort usually used in a teaching machine in that it was a non-programmed rote learning task.

4. Isolation Condition With Aural Feedback. This treatment was identical to the one just described, with the exception that the experimenter sat nearby and gave vocal feedback as soon as the S had announced his response. If the S gave the correct response, the experimenter would say, "right," "that's right," or "correct." If the response was incorrect, the experimenter would say nothing.

The Sample

Those who participated as subjects in the experiment were 180 elementary school children (91 males and 89 females). All Ss were naive as to the purposes of the study; however, one S was removed because she had some knowledge of German gained through a mother of German birth.

The participants were assigned to particular treatments on the basis of a reading score. Measures of reading ability, as well as intelligence measures were already available in the schools. Reading was selected for equating groups because: (1) performance on the task required some reading competence, and (2) reading ability and intelligence measures have been demonstrated to be highly correlated. The publishers of the Metropolitan Reading Tests report correlations between the Metropolitan Reading subtests and the Pintner General Ability Tests (4th grade--Pintner-Durost, Scale 2/5th, 6th grade--intermediate verbal) of .83 to .90. These correlations were acquired during the publisher's 1958 national standardization program. However, it should be pointed out that the learning task which was used was not one which would place high premium on intelligence as measured by such tests. Wechsler (1944, p. 4, 84) points out that intelligence is characterized by only normal memory ability, and that memory span for digits "correlates very poorly with all other tests of

intelligence." The learning task employed in this study was similar to digit memory span in that both involve rote acquisition.

Subjects were matched across all treatments where comparisons were to be made. For instance, as was mentioned in the introductory chapter, the investigator was interested in whether or not there was a learning advantage for the S at close proximity to the experimenter giving feedback. This meant that in order to test the proximity variable, subjects at greater distances had to be matched with those at lesser distances. Subjects A and D, B and C, 1 and 4, etc. of the eight member group were matched as closely as the available range of scores would permit. Then in order to compare direct reinforcement conditions against the vicarious condition and the isolated conditions, it was necessary to match Ss across these comparisons also.

Subjects

The subjects for any particular group were selected from a single grade (4th, 5th, or 6th) within a single school. All treatments were equally represented within each school-grade. Thus the subjects in isolation conditions 3 and 4 ended up with 44 and 40 Ss respectively instead of the intended 48.

Three elementary schools in Salt Lake City provided Ss for the experiment and one in Bountiful. The three in Salt Lake were Oquirrh, Forest, and Stewart and the one in Bountiful was the Boulton Elementary. The first of these is located in an obsolescent residential section of the downtown area. The second represents a middle class population, and the third is the teacher training laboratory school of the University of Utah. The fourth school, the Boulton Elementary of Bountiful, Utah, is in a suburban residential area.

Experimenter's Instructions

Instructions to those in a group of eight were as follows:

"This is a part of school; however, it resembles a game more than anything else. We will be learning to use words from a foreign language. To increase the fun, I'm not going to tell you what language it is until a few days later.

It's important that you pay careful attention, even though I may not be talking directly to you. Don't be concerned about your turn to answer the problem; instead, try to learn the meaning of the foreign word. Some people won't be asked for many answers (4/4 vicarious Ss were only given one word at the beginning of the first and second days; these words did not count in the analysis). We'd like everyone to learn all the words he can, but, or course, it's pretty hard to learn them all. (At this point a card was placed on a stand on the experimenter's desk.--see Figure 2).

"Now, let's practice. I'll call your name, and read the foreign word to you. You will then read the two English words aloud, and tell me which one you think means the same as the foreign word. I'll call on only one person at a time, and only he should answer.

"Okay, (S's name) your word is "Saugling;" read the two English words and tell me your choice."

When the subject had made his response to each item, the experimenter said, "right," "good," or "correct" unless the response was wrong, in which case the experimenter said nothing. After the feedback had been given, the card showing the problem was kept in view for four seconds in order to permit subjects to fixate the correct choice.

Isolates instructions were similar. Instead of the group instruction relative to making oral responses, they were told how to advance the machine, how to respond subvocally, and how to bring the correct answer into view. The subject was told to hesitate after seeing the answer, so that he would be able to remember it. He was then told to proceed on his own, and to try to connect the foreign word with the appropriate English

word. Condition IV was the same as Condition III except that the teacher sat by the pupil and said "right" or "correct" on appropriate occasions.

The Experimenter, The Data Recording

The experimenter who presented the learning task to the group was a young woman who had several years of teaching experience. On occasions, when she was presenting the material to be learned, the scientist in charge sat at the rear, or at the side of the room, in order to record the responses of the subjects. A separate record was kept of each subject's performance on particular items presented to him.

Another young woman directed the learning under Treatment 4 which involved isolation with aural feedback from the experimenter. This experimenter kept her own record of response performance in as inconspicuous a manner as possible. Right responses were indicated on a list by placing a dot in front of correct items.

The scientist in charge served as experimenter for the isolation Condition 3 subjects. After an initial explanation of how to proceed with the "teaching machine," no more contact was made with the subject unless the machine jammed.

The Task To Be Learned.

The learning task as previously described consisted of a German stimulus word which in the experiment was to be correctly associated with one of two English noun responses. Figure 2.02 shows the format of the task as it was presented by the teacher-experimenter to each subject.

The acquisition trials were conducted for three days: Monday, Tuesday, and Wednesday. On each of these days a set of twenty words was presented with four trials on each. The total of eighty presentations per day required approximately forty minutes to complete. A criterion

SAUGLING

BABY

GIFT

Fig. 2.02 Example of task presentation format. The top word is German meaning baby. Subject was to select either Baby or Gift as the equivalent of Saugling. On repeated presentations choices were alternated as to position, i.e., Gift appeared on top, Baby on bottom. Four trials on each card were allowed. Cards were 11 x 14 inches.

measure was not taken until Friday.

The task was presented on a series of cards similar to the one illustrated. The experimenter stood in front of the row of 8 pupils and held up a card. He then read the German word and the two English words. The pupil designated to respond to the particular card then guessed which answer was correct. The experimenter then either told him he was right or took up the next card and turned to the pupil who was to respond to it. On each day the experimenter went through the cards four times in two different orders.

Criterion Measures

A test of retention was given on Friday, This was four days after the first learning trials (Monday), and only two days after the last of the trials (Wednesday).

The test was intended to measure recognition skill. It presented the stimulus word and offered four choices for the response as shown on pages 1.13 - 1.15.

Results

Acquisition Learning Curves

Learning Condition 1 and 2 involved learning in a group of eight subjects: Half of these subjects (four)responded orally to the problems as presented on cards by the experimenter. For these Ss measures of learning were obtainable while acquisition was in progress. However, since the remaining four Ss (Condition 2) did not overtly respond, no measure of acquisition could be had from them during the three days of learning trials.

There were two isolated (learning alone) treatments each with four Ss per experimental set. In one of the treatments the participants worked alone. These Ss received reinforcing visual feedback from "teaching machines" after responding subvocally. Since responses were covert, no acquisition measures could be obtained from them while learning was in progress. The other isolated treatment involved a single S at a time giving oral responses to an experimenter who watched the S operate a "teaching machine." For this group a measure of learning was recorded by the attendant experimenter.

The experimental conditions provide acquisition learning curve data for Condition 1, direct oral responding in the group of eight, and for Condition 4, oral responding of isolated Ss in the presence of an experimenter. Comparable tasks were used on each of three days. These tasks have been designated A, B, and C. Each task involved twenty problems (a German noun stimulus, and two English noun response choices); thus, there was a total of sixty items. Each one of these was presented four times with the four presentations scattered through the eighty trials of

the series.

Acquisition learning curves are shown in Figures 2.03 and 2.04. The points plotted on the curves were obtained as follows: (1) the performances of all Ss were summed together separately for the first, for the second, for the third, and fourth presentations of specific items; (2) a mean was determined for each of the four presentations; (3) these means were then converted to a scale of 20, since 20 items constituted each day's task. To illustrate, there were 48 subjects involved in Condition 1. Each responded orally to five items per day. This resulted in 240 first trial subject-items. Had it happened that on the first trial there had been 120 correct responses, the mean performance would have been .5. Such a performance would be placed at ten on the vertical scale in Figure 2.03, since 20 items constitute one day's task. In reality, as seen in Table 2.01, the first presentation performance was in fact only slightly higher, or 10.25. Since there were two alternatives for each item in a twenty-item task, the first trial performance level of these S (approximately .5) is very near the mean chance expectancy. Six first trial measures (two groups--three days) are all consistent in this respect.

This is fairly good evidence that mean association values are near zero, but slightly and consistently positive. That is, taking the task as a whole, negative association values are practically equal to positive values. The learning subject chooses almost, but not quite, as many wrong responses as he chooses right ones.

No evidence is available from these curves to indicate the extent of differential association value when considering individual items. There may be great disparity, or the association values may be homogeneous.

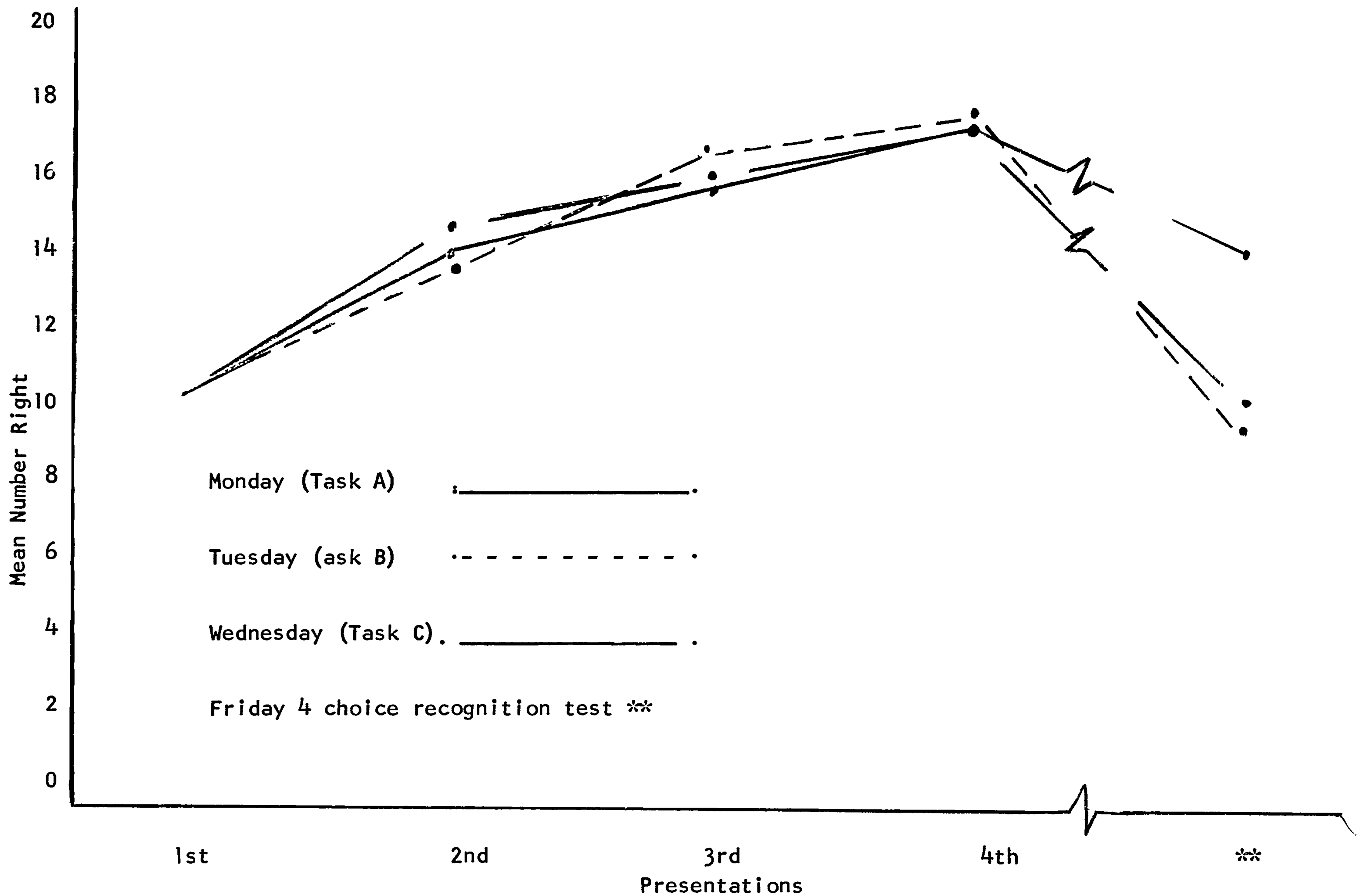


Figure 2.03 Mean performances for direct reinforcement subjects (oral respondents in group) on two-choice acquisition trials and delayed test.

TABLE 2.01

MEANS AND STANDARD DEVIATIONS FOR DIRECT REINFORCEMENT Ss IN THE GROUP SITUATION ON ACQUISITION TRIALS CONSISTING OF FOUR PRESENTATIONS FOR EACH STIMULUS WORD AND WHERE EACH TASK INCLUDES 20 WORDS SHARED BY FOUR Ss

	Presentations							
	1st		2nd		3rd		4th	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
MONDAY TASK A	10.25	2.22	14.0	2.48	15.75	2.1	17.17	2.33
TUESDAY TASK B	10.25	1.50	13.83	2.31	16.75	2.5	17.75	1.5
WEDNESDAY TASK C	10.25	1.89	14.58	2.4	15.92	2.18	17.17	1.82

Note: Twenty items having two alternatives were included in each task. The mean expectancy on the first presentation was chance, or 10 correct.

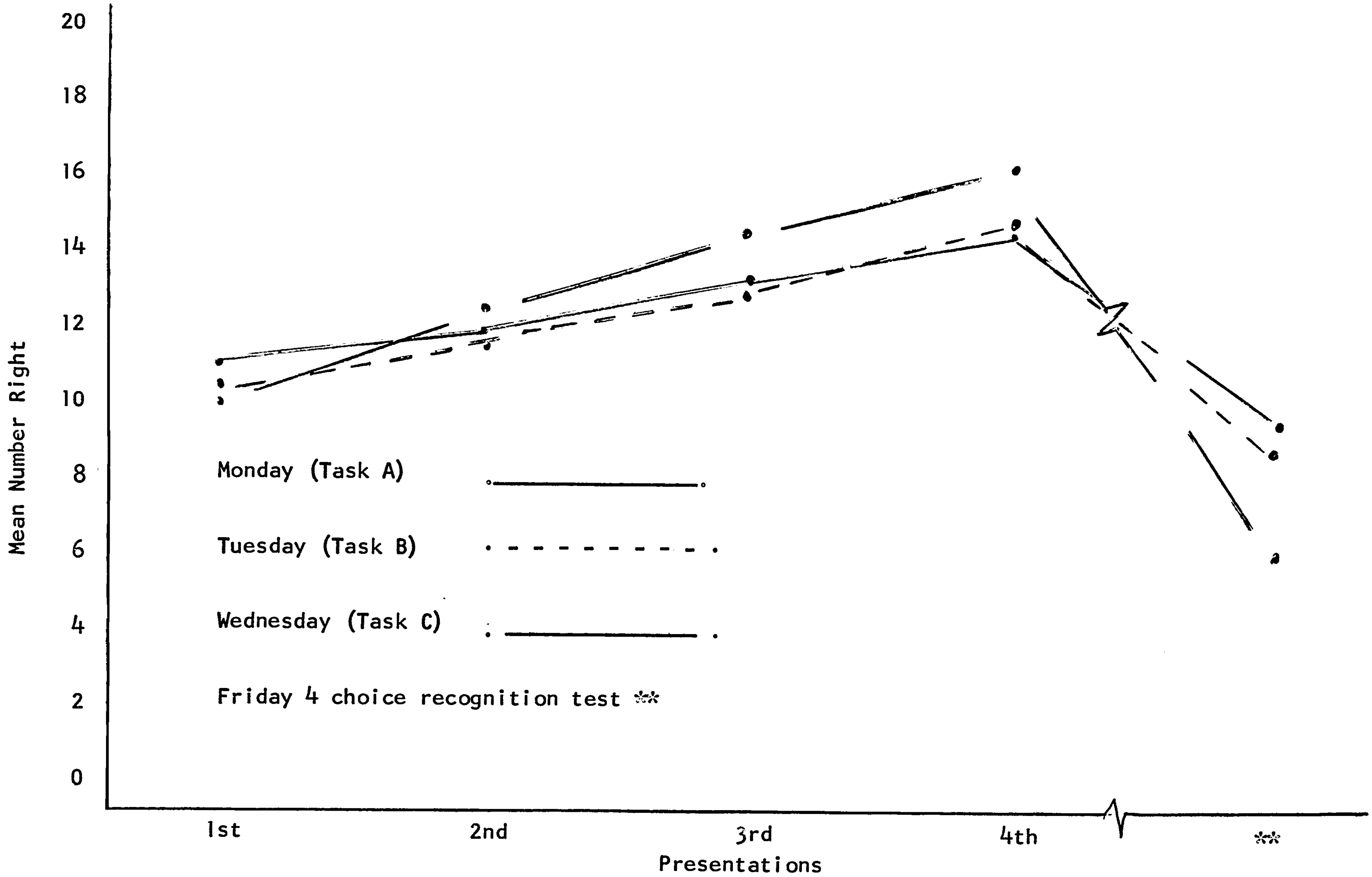


Figure 2.04 Mean performance for isolates (Condition 4 <u>S</u>s reinforced by teaching machine and with aural feedback from experimenter) on two-choice acquisition trials and on delayed test.

TABLE 2.02

MEANS AND STANDARD DEVIATIONS FOR CONDITION 4 (ISOLATION-TEACHING MACHINE AND AURAL FEEDBACK) ON ACQUISITION TRIALS CONSISTING OF FOUR PRESENTATIONS FOR EACH STIMULUS WORD AND WHERE EACH TASK INCLUDES 20 WORDS PRESENTED TO EACH S

	Presentations							
	1st		2nd		3rd		4th	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
MONDAY TASK A	11.38	2.86	12.25	2.59	13.55	2.95	14.93	3.56
TUESDAY TASK B	10.63	2.69	12.10	3.33	13.30	3.41	15.10	3.59
WEDNESDAY TASK C	10.35	3.31	12.63	2.68	14.9	2.22	16.65	2.63

Note: Twenty items having two alternatives were included in each task. The mean expectancy on the first presentation was chance, or 10 correct.

More information will be offered on this in the section on identification or copying behavior. Without sufficient evidence to show that individual items were equivalent in terms of relative association values it was necessary to use identical items on both sides of all comparisons. That is, all learning conditions involved the same task, item for item.

Looking at the curves for the direct reinforcement group (Figure 2.03), the reader will note a slightly negative acceleration. There is a more rapid rise at the second presentation than is the case for isolated Ss

(Figure 2.04). The advantage remains with the direct reinforcement group for all presentations, over all three days.

Acquisition for the isolation group was almost perfectly linear. In addition, the reader will note a nearly identical performance on Monday to that of Tuesday and Wednesday. In relation to the consistency of these learning trials the criterion performance measured on Friday is of notable interest. The graphic comparisons afford a ready grasp of the original learning and later forgetting pattern (Figure 2.04). When learning is measured by a delayed test, notice that primacy in learning holds a strong advantage over recency on this task. Monday's material is remembered fa better than is Tuesday's or Wednesday's (Sig. <.001, F = 82.7 with 2 and 468 d. f., see Table 2.03).

<u>Criterion Measures of Retention--Analysis of Variance of Treatments by Day of the Learning Trials</u>

In order to examine the major outcomes from the retention data collected with the recognition test on Friday, a four by three analysis of variance design was employed. Differences in the four learning conditions were compared on one dimension and days of the learning trials on the other.

Only the first ten experimental sets were used. This resulted in 40 Ss per learning condition being used. Table 2.03 summarizes the analysis of variance. As already reported in connection with the learning curves, the effect of days was highly significant. The first day's task was remembered better, at a highly significant level, than the following two days. The lesser ability to remember the more recent material was not quite so dramatic for Condition 3, isolation, as for the others. An interaction effect (Sig. beyond .01) resulted largely because of this.

TABLE 2.03a

MEAN TEST PERFORMANCE ON
FRIDAY FOR EACH DAY'S TASK

	1. Direct Learning	2. Vicarious Learning	3. Machine Feedback Only	4. Machine Feedback plus Teacher Feedback
Monday (20 items)	12.80	11.85	11.73	10.03
Tuesday (20 items)	8.40	8.03	9.18	9.43
Wednesday (20 items)	7.40	6.43	8.33	6.90
Total for entire task	28.59	26.31	29.22	26.36

TABLE 2.03b

ANALYSIS OF VARIANCE FOR LEARNING CONDITIONS
AND DAYS OF THE LEARNING TRIALS

Source of Variation	Sum of Squares	df	Mean Square	F	Required .05	Required .01
Between days	1554	2	777	82.7	2.99	4.60
Between learning conditions	92	3	31	3.3	2.68	3.95
Interaction	198	6	33	3.51	2.17	2.96
Within	4419	468	9.4			
Total	6263	479				

Isolation Condition 3, where the S learns the task as it is presented by a teaching machine, had the highest total performance. While this learning condition manifests the highest over-all performance, and the highest terminal performance (Task C, learned Wednesday), its learning was next to the poorest on Task A (learned Monday). Again, this evidences the significant interaction effect of days of learning and treatment conditions.

The learning conditions also had significantly differential effects. In order to determine where such differences lay, a Scheffe Test of Multiple Comparisons was carried out. The results are presented in Table 2.04.

TABLE 2.04

SCHEFFE TEST OF MULTIPLE COMPARISONS
FOR LEARNING CONDITIONS TABLE 2.03

Condition 1, consisting of four _S_s in a group of eight who responded orally in turn, had a total performance not significantly different from Condition 3 (isolation), the highest performing group. Had the learning trials been extended another day or two, this might not have been the case, for Condition 1 had a higher relative rate of performance decay over days than did Condition 3.

An inspection of the _Scheffe Test_ shows that Condition 3 (isolation), exceeds Condition 2 (4/4 vicarious) beyond the .01 level of significance. The mean for Condition 3 (isolation) was also reliably greater than that for Condition 4 (isolation with experimenter feedback) beyond $p = .01$. Condition 1 subjects were also reliably superior to Conditions 2 (4/4 vicarious), and 4 (isolated--with experimenter feedback) but at a level of significance less than .01. The only comparisons not significant were 2 and 4, and 1 and 3.

At this point, it is necessary to remind the reader that Condition 1 was a mixed direct and vicarious treatment. That is, two learning conditions were in effect for these _S_s. One-fourth of all items were experienced by each of the four _S_s directly. Each responded orally to particular items, in turn, and was reinforced with feedback by the experimenter for a correct response. For the remaining three-fourths of the items these subjects had to learn vicariously, that is, by listening to the exchange between the active _S_ and the experimenter. Tables 2.05 to 2.07 show the two performances in Condition 1 broken down as percentages according to the one-fourth--three-fourths learning modes. Direct responding in the group (1/4) was superior to all other modes of learning studied. In each measure it was as high as any other condition, or higher.

TABLE 2.05

PERCENT OF ITEMS CORRECT ON TASK "A" PRESENTED ON MONDAY, TESTED AS RECOGNITION ON FRIDAY*

CONDITION 1 (social) (N = 48)					
	PROXIMITY TO EXPERIMENTER				
	1	2	3	4	TOTAL
	Direct Aural Feedback (1/4 of total items)				
	64	73	69	57	66
	Vicarious Aural Feedback (3/4 of total items)				
	68	58	62	54	61
CONDITION 2 (social) (N = 48)					
	PROXIMITY TO EXPERIMENTER				
	A	B	C	D	
	Vicarious Aural Feedback (4/4 of total items)				
	55	56	56	67	59
CONDITION 3 (isolated) (N = 44)					
	Feedback: 1. visual by machine				59
CONDITION 4 (isolated) (N = 40)					
	Feedback: 1. visual by machine 2. aural from experimenter				50

*Based on a four-choice test making chance expectancy approximately 25 percent.

TABLE 2.06

PERCENT OF ITEMS CORRECT ON TASK "B" PRESENTED ON TUESDAY, TESTED AS RECOGNITION ON FRIDAY

CONDITION 1 (social) (N = 48)					
	PROXIMITY TO EXPERIMENTER 1	2	3	4	TOTAL
	Direct Aural Feedback (1/4 of total items)				
	44	50	53	42	47
	Vicarious Aural Feedback (3/4 of total items)				
	47	41	36	43	42
CONDITION 2 (social) (N = 48)					
	PROXIMITY TO EXPERIMENTER A	B	C	D	
	Vicarious Aural Feedback (4/4 of total items)				
	33	40	41	45	40
CONDITION 3 (isolated) (N = 44)					
	Feedback: 1. visual by machine				46
CONDITION 4 (isolated) (N = 40)					
	Feedback: 1. visual by machine 2. aural from experimenter				47

TABLE 2.07

PERCENT OF ITEMS CORRECT ON TASK "C" PRESENTED ON WEDNESDAY, TESTED AS RECOGNITION ON FRIDAY

CONDITION 1 (social) (N = 48)					
	PROXIMITY TO EXPERIMENTER 1	2	3	4	TOTAL
	Direct Aural Feedback (1/4 of total items)				
	49	53	47	50	50
	Vicarious Aural Feedback (3/4 of total items)				
	38	39	29	34	35
CONDITION 2 (social) (N = 48)					
	PROXIMITY TO EXPERIMENTER A	B	C	D	
	Vicarious Aural Feedback (4/4 of total items)				
	37	33	30	35	34
CONDITION 3 (isolated) (N =44)					
	Feedback: 1. visual by machine				42
CONDITION 4 (isolated) (N = 40)					
	Feedback: 1. visual by machine 2. aural from experimenter				35

The subjects who responded directly to 1/4 of the items were compared with respect to their knowledge of these items with the subjects who had observed these responses. This comparison was made by means of a t test for the scores on these items. The results are in harmony with the analysis of variance, showing a significantly better performance for direct participation than for vicarious participation ($t = 2.30$ with 47 d.f., one-tailed test, $p < .02$). This comparison was made prior to the analysis of variance and serves only to corroborate the finding already reported as significant there.

The Data from the Subjects Who Learned Vicariously

What has been called vicarious learning has been shown in the present study to be not as effective as learning by more direct means. The suggestion has commonly been made that the efficiency of learning varies according to the extent of the subject's involvement. An occasional response given orally in a group, as was the case for Condition 1, may have the capability of maintaining sufficient arousal in the respondent so that his learning efficiency for other items to which he does not respond is enhanced. In contrast, if he never makes an overt response, we might expect his involvement in the relevant learning activity to drop off.

A comparison was made between the learning on the items to which the subjects responded directly and the learning for these same subjects on the items to which they did not respond directly.

A sign test (Dixon and Massey, 1951) of the difference between the sets of data offers evidence of greater performance decline over the three days for items vicariously experienced than for those overtly responded to. That is, what is commonly considered to be the proactive inhibition effect

was greater for vicarious learning than for learning by direct experience. Table 2.08 shows how the 48 subjects performed on items directly experienced and on items vicariously experienced. The data originated from the retention test given on Fridays of each week of the experimentation. A plus sign indicates that in that experimental group on the day specified, direct reinforcement subjects learned proportionately more items by direct experience than they did by vicarious experience. A minus sign occurs where Ss learned more vicarious items than direct ones. A majority of measures favored the directly experienced items (p .01). It should be borne in mind that this is not a weighted test. The weightings which have no power in the significance test may be seen as summed numerical differences favoring positive (direct), or negative (vicarious) treatments. The mean difference favors the direct condition on Monday's task. The advantage favoring direct responding increases on Tuesday, and again on Wednesday, suggesting the hypothesis that attention (or whatever factor is operating) falls off faster for vicarious Ss than for direct Ss.

A simple analysis of variance was carried out to test this hypothesis. Difference values (direct minus vicarious) were used as the basis for analysis.

As can be seen in Table 2.09, the trend failed to show significance at the .05 level.

The trend suggesting more rapid decay of learning for vicarious Ss than for direct may be of sufficient interest to warrant an extension of this design to include an additional day or two of learning trials. This would permit a re-examination of the differential performance decay with the possibility of a significant trend showing up.

TABLE 2.08

SIGN TEST OF DIRECT 1/4 ITEMS VERSUS VICARIOUS 3/4 ITEMS*
NUMERICAL VALUES SHOW THE DIFFERENTIAL ADVANTAGE (X-Y)
HELD BY THE CONDITION OF THAT SIGN

Task		A - Monday	B - Tuesday	C - Wednesday
Experimental Set	4	+1	1-	+4
	5	+1	+4	+1
	6	+3	+3	+7
	7	+2	+4	+1
	8	+4	+8	+2
	9	3-	1-	0
	10	+2	1-	+4
	11	2-	3-	0
	12	2-	0	+6
	13	1-	+1	0
	14	+2	1-	+8
	15	+4	+1	+4
+ n		8	6	9
- n		4	5	0
+ Σd		19	21	33
- Σd		8	7	0
+ $\bar{X}d$		2.38	3.5	4.12
- $\bar{X}d$		2.0	1.4	0

*A plus indicates an advantage for direct, a minus for vicarious.

TABLE 2.09

TREND ANALYSIS: ANALYSIS OF VARIANCE OF DIFFERENCE SCORES

(1/4 DIRECT MINUS 3/4 VICARIOUS) BETWEEN DAYS

Source	Sums of Squares	df	Variance Estimate	*F*	F required at .05
Between days	6.70	2	3.35	2.11	2.99
Within days	224.66	141	1.59		
Total	231.36	143			

<u>100 Percent Vicarious Versus Partial Vicarious</u>.

An attempt was made to demonstrate a difference in learning for subcondition 1b (3/4 vicarious) when compared with Condition 2 (4/4 vicarious) on the items held in common by these treatments. A significant difference favoring 1b would possibly indicate a general arousing capability for the condition where <u>S</u>s give intermittent responses. This arousing state might then be expected to manifest itself by higher performance on the 3/4 vicarious items. The contrast would be against the relatively deprived activity state of subjects in the 4/4 vicarious condition. Again, the learning for both conditions was by the vicarious mode, and the measurement made only on those items held in common by both sides of the comparison.

The percentage Tables 2.05-2.07 indicate a higher performance level for 3/4 than for 4/4; i.e., Task A percentage right for 3/4 vicarious items was 61% and 4/4 items was 58%. Five of the six measures in the

tables favor 3/4. A t-test of correlated means, having 47 degrees of freedom was not significant, however. It must be assumed, therefore, until a more sensitive measure can be taken, that the difference is not reproducible.

Proximity to the Experimenter, and to the Responding Subject.

One of the questions asked in this study was whether or not there might be a difference in learning for near and distant seating positions in relation to the "teacher"-experimenter who was positioned near the center of the row. The hypothesis was that the closer proximity would result in greater learning. An inspection of the data for Condition 1 (1/4 direct and 3/4 vicarious) indicated no substantial proximity position effect. Such was not the case with the 4/4 vicarious data. The recognition data appeared to favor seating position D over C and B, and they inturn over seating position A. The individuals at greater distances from the center of the row had higher performances, with the individual at the end showing up best.

A four by three analysis of variance compared positions and days of the learning trials. The results of this analysis are shown in Table 2.10.

Indications are either that:

1. The apparent proximity effect for Task A and B is simply a chance variation of the data, or,

2. The effect is real but temporary; otherwise, Task C variance is simply a chance fluctuation from the "true" population value.

Table 2.10 shows that the proximity effect across three learning tasks combined was not significant, though it does not fall far short of the

5 percent level. The interaction term was not significant.

TABLE 2.10

ANALYSIS OF VARIANCE FOR PROXIMITY POSITIONS*

AND DAYS OF THE LEARNING TRIALS

Source of Variation	Sums of Squares	df	Mean Square	*F*	Required .05	.01
Between days of the learning trials	638	2	319	43.7	3.09	4.82
Between vicarious positions	55	3	18.33	2.5	2.70	3.98
Interaction	45	6	7.5	1.03	2.19	2.99
Within	961	132	7.3			
Total	1763	143				

*Positions in Condition 2, 4/4 vicarious.

Vicarious Ss proximity to the source of responding was also examined. There is the circumstance where a Subject "D" is seated at the end of a row of eight subjects. In this event the question may be asked: Does he (Subject D) learn items better that are responded to by a Subject (3) who is seated immediately next to him, or does he more easily learn the responses of a distant Subject (4) who is at the opposite end of the row? A one-tailed hypothesis held that items presented to S (3) would be learned better by vicarious S (D) than those items presented to S (4). This was expected because of a supposed greater arousing potential for a near by stimulus-response exchange. While differences were in the expected

direction, they failed to reach an acceptable level of significance. A t-test of the differences between means for correlated data was 1.24 with 35 degrees of freedom. At the .05 level, a t of 1.697 is required.

Discussion

Some significant differences were found among conditions of learning considered here. These may perhaps be very important differences. When performances for the three acquisition days were combined on the criterion test, the 4/4 vicarious scores totaled to 82 percent of the 1/4 direct scores. The difference between these two data groups was the greatest of any of the comparisons made. In appraising these circumstances one needs to keep in mind at least two relevant points: (1) a trend, which in this study fell short of significance at the .05 level, suggested that the difference separating direct and vicarious experience increased as time over the learning trials increased, and (2) the prevailing experimental conditions were distinctly ad hoc, suggesting that the attention level of vicarious Ss was probably higher than might be expected under circumstances where adaptation to the environs had occurred.

Statistically significant differences between the four treatments are indicated below by symbols (x) and (y). The same symbol appearing before two learning conditions indicates that the mean performances in the two conditions do not differ at the .05 level. Different symbols indicate significant differences of at least .05.

(x) Condition 1. Two conditions operating:

(a) 1/4 of the items were directly responded to.

(b) 3/4 of the items vicariously experienced.

(x) Condition 3. Isolated from other Ss. Working alone at a "teaching machine" from which visual feedback was afforded.

(y) Condition 2 (4/4 vicarious). All items are experienced by observing the Ss in Treatment 1 respond.

(y) Condition 4. Isolated and working under the same conditions as Treatment 3 except that an experimenter was present to give aural feedback as an additional condition.

Condition 3 (working alone) had the highest over-all performance. This was so in spite of a comparatively poor first day's performance. Both Conditions 3 and 4 (isolated conditions) had poor beginnings when compared with Conditions 1 and 2 (group conditions). By the third day of learning trials, however, the Condition 3 isolates were the best performers and sufficiently so to give this treatment the highest over-all performance (though not significantly different from Condition 1). These outcomes give rise to the following questions:

1. In the isolated conditions the Ss had control over presentation of their own items by use of a "teaching machine." The to-be-learned material was adapted to "teaching machines" with the format of each item being comparable to the card presentation shown in Figure 2.02 cards used to present the task to groups). The question arises as to whether or not poor initial performance could be accounted for in terms of time needed for adaptation to the unfamiliar mechanical device. Attention is

expended in part on various of the unfamiliar elements of the new environment. When there are distracting conditions, one might speculate that relevant attention is diminished to the extent that distractors use up the channel capacity of the perceptual system. However, Azrin (1958) has shown that noise inhibits discrimination, if the discrimination had been originally learned under quiet conditions. Conversely, when the discrimination behavior had been learned in the presence of noise, quiet was then disrupting. Habituation to the environment apparently can permit more effective selection of stimuli to be admitted to the perceptual system.

2. The reader will recall that Condition 1 actually involved two learning conditions (1/4 direct responding, and 3/4 vicarious responding). Even though Condition 3 (isolation) had the highest total performance, if the learning of specific items is examined, then the items directly responded to and directly reinforced in the group situation were more effective. Even on the third day of learning trials when isolated Condition 3 Ss had their best day, learning was still inferior to the direct oral response and reinforcement mode in the group situation. When 1/4 direct and 3/4 vicarious items were combined in Condition 1, the level of performance was slightly less, and not significantly different from, isolated Condition 3. Could one perhaps account for the generally superior learning

under the oral response mode in the following way:

a. When a S's name was called, in his turn, indicating that he should make the oral response, such a response required an obligatory orientation to the task, at least, attention had to center here sufficiently long for the subject to vocalize a response. For the other Ss, vicariously experienced items (observation of another's responding) entailed no such necessity for an arousal posture appropriate to the learning task. The subject was "free" to engage in other activity or to dream, if so inclined. The fact that the greatest learning, of all of the conditions of learning studied, occurred as a consequence of oral responding, and the poorest learning occurred in the passive situation of sitting in the presence of those making oral responses is suggestive that the study of attention processes lead to an understanding of the nature of certain learning mechanisms.

b. In addition, it is quite conceivable also that intermittent direct responding is capable of elevating arousal level. This higher general activity level might then increase the subjects sensitivity to task stimuli.

Reference

Azrin, N. H., Some effects of noise on human behavior. _J. exp. Anal. Behav._ 1958, 1, 183-200.

CHAPTER III

LEARNING AS A CONSEQUENCE OF THE LEARNER'S TASK INVOLVEMENT UNDER DIFFERENT CONDITIONS OF FEEDBACK

In the study reported in the previous chapter subjects learning in groups of eight were exposed to two main learning conditions. Four of the subjects learned by a recitation procedure, interacting with the experimenter who functioned as a teacher. These subjects were referred to as the direct subjects. The remaining four were able to learn the task only by observing the performance of the direct subjects and their interaction with the experimenter. The latter four subjects were referred to as observer subjects and, according to current custom, their learning was designated as vicarious learning. The task involved in the study was the acquisition of German vocabulary. The experimenter-teacher presented the German words, one at a time, on large cards accompanied by two English words one of which was the equivalent of the German word. The words were printed in large letters clearly visible to the entire group. After a direct subject, designated by the experimenter, selected one of the English words as the equivalent of the German word, he was either told by the experimenter "That is right," when it was correct, or was told nothing when the response was incorrect. Although the observer subjects were provided with an equal amount of information when the direct subject responded correctly as when he responded incorrectly, the data suggested that they learned better when they observed a correct response than when they observed an incorrect one. This led to the hypothesis that the specific nature of feedback provided by the experimenter in such a situation would be an important factor in determining the learning of the observer subjects. This is the central problem on which the present study focuses.

While little research has been undertaken on learning in situations which simulate those of the classroom, a considerable amount of information

has been obtained on the relationship of feedback to learning under conditions where only one learner is involved at a time.

Most of the studies have been concerned with the effect of the experimenter saying right or wrong or nothing or combinations of these. The results of such research are difficult to fit together into a consistent pattern bacause of the varied nature of the subjects used.

Buss, who has conducted a series of researches with various associates (Buss, Wiener, and Buss, 1954; Buss, Braden, Orgel and Buss, 1956; Buss and Buss, 1956; and Ferguson and Buss, 1959), have generally found that the combination of no comment (N) for correct responses and "wrong" (W) for incorrect, or "right" (R) for correct responses and "wrong" (W) for incorrect, produces more rapid learning than the combination of R-N. Buchwald (1959a, 1959b) has taken the position that N acquires positive reinforcing properties when given in the combination N-W, and negative reinforcing properties in the combination N-R. Meyer and Seidman (1960) found that no comment appeared to have reinforcing properties with an 8-9 year old group but not with a 4-5 year old group. The younger, pre-kindergarten group seemed unable to utilize the information provided by silence on the part of the experimenter when the response was correct. Clearly, providing the best form of feedback is a relatively complex problem even in situations involving only one learner. The present study explores the relative effectiveness of different forms of feedback in a simulated classroom situation in which there are pupils who interact with the teacher and pupils who learn by observing the interaction.

Method

Experimental Design

In the present study, as in the previous one, a simulated classroom situation formed the context of the study. Eight "pupil"-subjects sat in a row facing the experimenter who functioned in the teacher role. The experimenter interacted with the odd numbered subjects, but not with the even numbered ones, except on a single demonstration trial when he interacted with all Ss. The experimenter presented a German word together with two English words which were printed in large letters on an 11" x 14" card. When presenting a card, the experimenter turned to a designated subject and said,"Your word is ________" (German word). The designated subject then read both English words and selected one as the correct translation. After the subject had made his selection, he received knowledge of results according to the condition to which his group had been assigned. The four feedback conditions were as follows:

Condition 1

a. S correct - "That's right."

b. S incorrect - "No, that's wrong."

Condition 2

a. S correct - E said nothing when S made correct response.

b. S incorrect - "No, that's wrong."

Condition 3

a. S correct - "That's right."

b. S incorrect - "No, that's wrong, (German word) means (English word)."

Condition 4

a. S correct - E said nothing when S made correct response.

b. S incorrect - "No, that's wrong, (German word) means (English word.)"

Nine groups of eight subjects each were assigned to each one of the conditions and in each group four of the subjects, the direct learners, interacted with the experimenter, while four, the observer subjects, learned the task by observing the interaction. The work was undertaken in an empty room of a school.

Subjects

The subjects (N = 288) were fourth, fifth and sixth graders drawn from three public elementary schools in Salt Lake City, Utah. Two of the schools were located in an older residential section of the city, and the third school was situated in a newly developed suburb.

Task

The task in the present study consisted of learning to match sixty German words with their English equivalents. This learning occurred over three consecutive days beginning Monday. The design of the task has been fully described in the previous study already referred to.

Procedure

Three learning sessions for each group took place on Monday, Tuesday and Wednesday mornings. Sessions lasted between twenty-five and thirty minutes. Usually three experimental groups were seen per week. On Friday a recognition test was given to that week's subjects to measure the amount of word-learning which had taken place.

On the basis of either reading scores, (obtained from the Metropolitan Achievement Test or the Science Research Associates Battery), or an intelligence quotient (from the Pintner General Abilities Test) eight, sixteen, or twenty-four subjects were selected from each of the fourth, fifth and sixth grade classes available in a school. Reading scores or I.Q.'s were ranked for each class, and subjects chosen counting up and down from the mean score. Thus experimental groups chosen from a class were kept as homogeneous as possible with respect to reading grade placement or I.Q. of the group's members. If a chosen subject was absent on Monday, another class member was substituted, keeping the reading score, or I.Q. as equivalent as possible to those of the rest of the group. If a subject was absent on a Tuesday or Wednesday, a substitute from his class was chosen for the learning days, but only the data of the original subject was used in the analysis. Children with any previous knowledge of German were eliminated from the sample. In one or two cases, it was discovered after the experiment had been run that a child with a German background had been used as a subject. In those cases, the subjects' data was not used in the analysis.

After assigning subjects to seats in the row of chairs, which was six feet in front of E, the experimenter gave the group the following instructions:

> "Although this is part of school, it is more like a game than anything else. What we do here won't have any affect on the rest of your grades. Only Miss H (the second E, who acted as recorder and observer) and I will know what you do here. You will be learning some words from a foreign language, but to

make it more fun, I'm not going to tell you what language you're learning until we're all through working in your school.

It is important that you pay very careful attention, even though I may not be talking directly to you. Don't worry about your turn to answer. Instead, try to learn the meaning of the foreign word. Some people won't be asked for many answers, but we'd like everyone to learn all the words he can. You'll be learning words Monday, Tuesday and Wednesday. On Thursday we won't come at all. But on Friday we'll come again and see how well you remember the words you learned. So try to remember as many words as you can.

(At this point E put a task card on a stand which was placed on a table in front of her.)

"Now let's practice. I'll call somebody's name, and then I'll read the foreign word aloud. The person whose name I call, will then read aloud the two English words which are underneath the foreign word. Then that person will guess which one of those two English words means the same thing as the foreign word. The person I call on will know if he's right or wrong because I will tell him so, (in the case of Conditions 1 and 3). (When using Conditions 2 and 4, E said, "If ___(S)___ makes a wrong guess, I'll tell him so, but if he's correct, I won't say anything at all.") I'll call on only one person at a time, and only he should answer. Let's have a few practice tries. O.K. (one of observer S's names), your word is 'Stirn.' Read the two English

words aloud and then tell me which one you think means the same thing as the foreign word. At first it will be like a guessing game, but after a while you'll be more sure of your answer." Miss H. is the score keeper for this game. (S read the words, E gave appropriate feedback, and E gave a practice word to each of the three other observer subjects.)

"All right, are there any questions? Now let's start with the first word that's really part of the game."

The other features of the procedure were the same as those involved in the study reported in the previous chapter. Twenty new words were learned each day on Monday, Tuesday and Wednesday, and the retention test was administered on Friday. The task was presented on the same cards as were previously used. In administering the test, the experimenter read each German word aloud and then paused about ten seconds for the subjects to respond.

Results

Effects of Feedback Mode, Subject Involvement, and Days of Learning

An analysis of variance based on the recognition test data from 288 Ss is provided in Table 3.01. This test data, obtained on Fridays, provided separate scores for the learning which had occurred the previous Monday, Tuesday and Wednesday. The data also provided separate measures for those Ss involved directly in learning German vocabulary and for those who learned by observation only. In addition, equal numbers of subjects learned under each of four feedback (knowledge

Table 3.01

ANALYSIS OF VARIANCE FOR FEEDBACK CONDITIONS

DAYS OF LEARNING, AND EXTENT OF SUBJECT INVOLVEMENT *

Source of Variation	df	Mean Square	F	Required .05	Required .001
Between Feedback conditions (L)	3	65.3	5.8	2.6	5.4
Between Direct (C) learners and observer learners	1	228.0	20.3	3.8	10.8
Between days (R)	2	1073.5	99.1	3.0	6.9
R x C	2	38.0	3.4	3.0	6.9
R x L	6	10.0			
C x L	3	3.0			
R x C x L	6	6.3			
Within	840	11.3			
Total	863				

*For basic data refer to Table 3.03

of results) conditions. It is apparent from Table 3.01 that all main effects were highly significant.

A Scheffe Test of Multiple Comparisons was used to assess differences between individual feedback conditions. These comparisons are given in Table 3.02. A relationship seems to exist between information content of the feedback condition and extent of learning. For example, under Conditions three and four, which were the most information-laden of the four feedback conditions, highest performance on the criterion test occurred. Learning under feedback Condition four was slightly, but not significantly, better than under Condition three.

Under the latter two conditions not only was a wrong response indicated, but the stimulus word and appropriate response were repeated by E in close temporal contiguity. This circumstance provided redundant information, since, in a binary choice situation saying, "No, that's wrong" gives complete information concerning the correct response choice. It should be noted that the correct response was the last phrase heard by the subjects under feedback Conditions three and four.

Condition three provided, in addition to the feedback given in Condition four, the utterance, "That's right," given by E when S responded correctly. This additional phrase, however, did not seem to add to the effectiveness of the feedback.

Condition one differed from condition three only in that it failed to add the phrase "____________________ means ______________." According to the Scheffe test learning under Condition three was not significantly more effective than under Condition one.

Table 3.02

SCHEFFE TEST OF MULTIPLE COMPARISONS FOR

LEARNING PERFORMANCE ACCORDING

TO FEEDBACK CONDITIONS *

Source of variation	F^a	P
A.		
4 - No, that's wrong, ___means___	46.31	.001
versus		
2 - No, that's wrong		
B.		
3 - That's right (and) No, that's wrong ___means___	29.85	.001
versus		
2 - No, that's wrong		
C.		
1 - That's right (and) That's wrong	15.99	.01
versus		
2 - No, that's wrong		
D.		
1 - That's right (and) That's wrong	7.88	.05
versus		
4 - No, that's wrong, ___means___		
All other comparisons not significant		

*For basic data refer to Table 3.03

It is obvious that feedback Condition 2, ("That's wrong,") was significantly inferior to all other conditions studied. No redundant information was provided by the experimenter and perhaps of consequence is the fact that "No, that's wrong," emphasizes a wrong response; thus increasing its availability for some Ss.

Primacy in learning was superior to recency under all conditions studied (P<.001). That is, the words learned on Monday were, under all feedback conditions, better retained on the Friday test than those learned on Tuesday or Wednesday. That which was learned on Tuesday was also better retained than the learning accomplished Wednesday.

Subjects directly involved in learning the task (odd-numbered Ss) were thus involved for only one-fourth of the task items while they learned as observers during the presentation of three-fourths of the items. This is to say that each of the directly involved subjects received only five of the twenty items presented on Monday, on Tuesday and on Wednesday directly. The remaining fifteen items could be learned each day by observing the interaction of other Ss with the experimenter. These two different conditions for learning were combined (1/4 + 3/4) in the analysis of variance in Table 3.01.

Since the learning for direct Ss was measured separately on those items learned directly and those learned by observation, these performances together with those for Ss whose total learning experience was by observation have been represented as percentages of the task learned and are provided in Table 3.03. In the previous study a knowledge-of-results condition "That's right" was employed in an experimental paradigm similar to the one used here. The percentage values from the earlier

Table 3.03

PERCENTAGE OF THE TASK RETAINED AT THE TIME OF THE FRIDAY TEST

		Direct subjects 1/4 items direct	Direct subjects 3/4 items by observation	Observer subjects all items by observation
Monday's task (20 items)				
Feedback	1	68	71	66
Condition	2	62	63	63
	3	69	65	67
	4	76	69	71
	[b]5	72	62	60
Tuesday's task (20 items)				
Feedback	1	54	51	48
Condition	2	61	49	46
	3	66	59	50
	4	61	54	51
	[b]5	50	42	39

Table 3.03 (Continued)

		Direct subjects 1/4 items direct	Direct subjects 3/4 items by observation	Observer subjects all items by observation
Wednesday's task (20 items)				
Feedback	1	65	51	46
Condition	2	58	45	43
	3	63	52	45
	4	68	52	46
	[b]5	52	35	33
Total task (60 items)				
Feedback	1	62	58	53
Condition	2	60	52	51
	3	66	59	54
	4	68	58	56

[a]N=72 Under each feedback condition
[b]Feedback condition - "That's right" from earlier study

study are included in Table 3.03 as feedback Condition 5. While Ss in the two investigations were from different elementary schools, with a different person, but of the same sex serving as the experimenter, other conditions such as grade in school, sex of the subject, and reading test scores appeared to be comparable. The "That's right" feedback mode resulted in less learning than any of those from the present study.

Retention of the task was measured again after the passing of about five months and is given in Table 3.04 as percent retained. This measure was acquired at only one of the schools where the sample was originally drawn. The retest was administered to 176 Ss, but these were not equally distributed across feedback conditions as may be observed in Table 3.04. It is worth noting that the proactive inhibition effect (advantage for primacy over recency) was still apparent, that is, Monday's task was better retained than Tuesday's, etc. In general, it appears that the four reinforcing conditions retained approximately the same relative order of effectiveness, with Conditions 3 and 4 being high and Conditions 2 showing poorest retention.

Interactions.. As may be noted in Table 3.01, an interaction between extent of subject involvement in learning and days of learning was significant at the .05 level. This interaction does not take into account the fact that the direct subjects learned under two conditions and, when these two conditions are separated, an additional component is added to the interaction. This is illustrated in Figure 3.01. Variance due to other interactions was in each case not significant. These included feedback-mode conditions by days, feedback-mode by extent of involvement, and the triple interaction of all main effects.

Table 3.04

PERCENTAGE OF THE TASK RETAINED AFTER

APPROXIMATELY FIVE MONTHS

		Direct subjects 1/4 items direct	Direct subjects 3/4 items by observation	Observer subjects all items by observation
Monday's task (20 items)				
Feedback 1	(N=16)	65	45	55
Condition 2	(N=16)	43	49	47
3	(N=72)	57	55	50
4	(N=72)	56	56	50
Tuesday's task (20 items)				
Feedback 1		40	38	41
Condition 2		35	33	33
3		48	41	40
4		45	38	41
Wednesday's task (20 items)				
Feedback 1		25	31	37
Condition 2		38	38	30
3		44	32	33
4		41	33	34

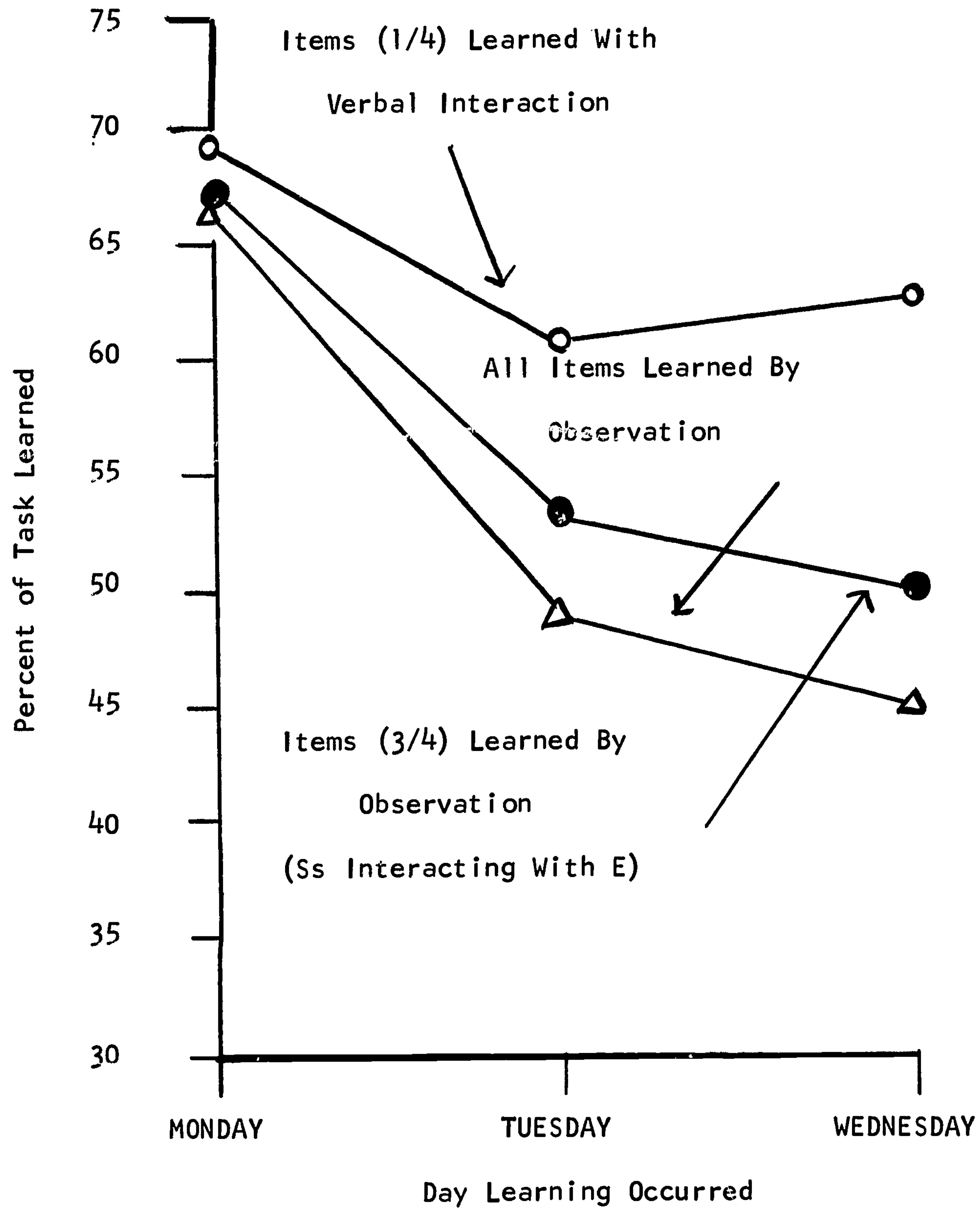


Fig. 3.01 Interaction between extent of subject involvement and days when learning occurred.

Discussion

Each one of the four learning conditions described provided all of the information necessary for the acquisition of correct responses, but the feedback differed in the amount of redundancy provided. The situations thus differed from those studied by Buss, Buchwald and others for they provided different forms of feedback giving equal information. The amount of redundancy is related to the degree to which the task is learned - with greater redundancy favoring learning. Of interest is the fact that those forms of feedback, in which the last item of information transmitted was the correct response, were significantly more effective than those in which other information was the last transmitted. This finding is in contrast with studies of R, N and W as reinforcers in which the combination R-N was generally proven to be less effective than the other two combinations.

Another point of interest raised by the study is that the subjects who interacted with the experimenter performed better not only on the items on which they interacted but also on the items which they learned by observation. The data suggest the interpretation that the direct interaction procedure raises the level of arousal of the direct subjects which, in turn, influences acquisition on the items which they learn by observation.

References

Buchwald, A. M., Experimental alteration in the effectiveness of verbal reinforcement combinations. J. exp. Psychol., 1959, 57, 351-361 (a).

Buchwald, A. M., Extinction after acquisition under different verbal reinforcement combinations. J. exp. Psychol., 1959, 57, 43-48 (b).

Buss, A. H., Braden, W., Orgel, A., & Buss, E., Acquisition and extinction with different verbal reinforcement combinations. J. exp. Psychol., 1956, 52, 288-295.

Buss, A. H., & Buss, E., The effect of verbal reinforcement combinations on conceptual learning. J. exp. Psychol., 1956, 52, 283-287.

Buss, A. H., Wiener, M., & Buss, E., Stimulus generalization as a function of verbal reinforcement combinations. J. exp. Psychol., 1954, 48, 433-436.

Ferguson, Elsie L., & Buss, A. H., Supplementary report: acquisition, extinction and counterconditioning with different verbal reinforcement combinations. J. exp. Psychol., 1959, 58, 94-95.

Meyer, W. J., & Seidman, S. B., Age differences in the different reinforcement combinations on the acquisition and extinction of a simple learning problem. Child Develpm., 1960, 31, 419-429.

Van Wagenen, R. K., & Travers, R. M. W., Some factors related to learning under simulated classroom conditions. J. educ. Psychol., in press.

CHAPTER IV

RELATIONSHIP OF LEARNING TO ATTENTION IN SIMULATED CLASSROOM SITUATIONS

While teachers and psychologists agree that the degree of acquisition of a skill is highly dependent upon the learner maintaining an orientation towards the task, there is considerably less agreement on what constitutes evidence that the learner is task oriented. In a typical verbal transaction in the classroom between teacher and pupils, such as occurs during recitation periods or during class discussions, teachers typically demand that each of the pupils direct his gaze towards the teacher or the discussant despite the fact that the information is being transmitted through the ear. Whether this required behavior facilitates the acquisition of knowledge is an open question. One can understand that when the information to be acquired is to be transmitted through the eyes that the direction of the gaze might be an important matter. However, when the information is transmitted through the auditory channel such a visual orientation would appear to be quite unnecessary. The data derived from the previous experiment permits the investigation of problems such as these since under two of the conditions information had to be obtained through the visual channel but under two conditions all of the necessary information was provided through the auditory.

In accordance with custom, the term <u>attention</u> will be used here to refer to what was described in the previous paragraph as <u>task orientation</u>. While the latter term refers to the matter of whether conditions of posture and conditions within the nervous system are set so that information from a particular source can be most readily received, only those aspects which involve posture can be readily observed. The assumption of many teachers is that the postural components are highly correlated with the observable components of attending, no assumption is made that these are correlated with

the components that cannot be observed directly.

Method

The Learning Task. The experimental arrangements were those described in Chapter 3. The task consisted of the learning of 60 German vocabulary words by members of an eight-member group. There were two primary conditions for learning: By direct verbal interaction with the experimenter and through obersvation or vicarious experience (observation of the direct respondant). In the case of direct learning, E held a card (11" x 14") providing an item, read the German word from the card, and called the name of a subject, who was to respond orally. The S then read aloud two possible English equivalents from the card, made a choice and received knowledge of results from E. Observational learning occurred as the remaining seven subjects observed the interaction of the direct respondent and experimenter.

The procedure followed by the experimenter in presenting the task was approximately as follows: E placed one of the cards (11" x 14") on a stand positioned on a large table before the Ss. She then addressed a subject by name, saying for example, "John, this word is Baum." In response, the subject read both English alternatives as the equivalent of Baum from the card and made a choice. The S might have responded "tree," the correct choice, or he could have said "corn," the incorrect response. Knowledge of results was provided from one of four categories:

Condition I
(nothing said when S was correct)
"No, that's wrong"

Condition II
"That's right"

"That's wrong"

Condition III
"That's right"

"No, that's wrong, ______ means ______."

Condition IV
(Nothing said when S was correct)
"No, that's wrong, ______ means ______."

The point to note is that the first two forms of feedback require the learners to direct their eyes towards the visually presented material, while in the case of the last two forms of feedback, the information can be obtained entirely by listening.

Design and Procedures

Thirty-six eight member groups were involved in the study. The subjects were the same as those in the previous study. Each group of eight was seated in a straight line facing the experimenter who stood about seven feet distant from the center of the row. Half of the subjects, the odd-numbered ones across the row were designated as direct subjects, that is, each one gave an oral response to one of the items. While one responded, the remaining three observed and attempted to learn the item. For those subjects that responded, one-fourth of the items were learned by verbal interaction and three-fourths by observation. Four Ss in the even numbered positions made no overt response at any time while the task was presented except on an initial practice trial. These Ss were referred to as observer subjects or as vicarious Ss. They depended entirely on observation for learning. Due to the fact that direct responding required an obligatory orientation to the task (the S could not escape looking, reading

and responding), manifestations of attention could be recorded only when a subject was involved as an observer; thus while one was directly involved, any one of seven Ss could be observed for evidence of relevant attention. Seated at the side of the room was an observing experimenter who, according to a schedule designating which S was to be observed, made the following objective recordings: (a) initial attention; credited as positive if the observed S was looking at the card for the entire interval beginning when E called the name of a direct subject until she had read the German word (a rapid glance away but returning again did not negate the recording), (b) concluding attention; was recorded positive if the S looked continuously from the time the direct S spoke his choice of response until E put the card down (a total of about five seconds). The objective assesment of attention was recorded on an observation schedule which also include the information as to the correctness of the direct S's response to each item each time it was presented.

At the conclusion of each days trials , the experimenter who had administered the task recorded her subjective estimation of which two of the eight Ss in the group had attended best and which two had attended least.

Learning trials occurred on three days: Monday, Tuesday, and Wednesday. On each of these days, 20 items were introduced with four trials on each. The four presentations of an item were programmed intermittently throughout the total of 80 item presentations per day so that it could not be anticipated which item would occur next in the series.

Test of Learning. On a separate Friday for each experimental replication, a four-choice recognition test was given to that week's subjects. Included in the four choices were the right response, the wrong one which

had appeared on the card during the trials, and two distractors selected from a total pool of response words.

Subjects. Subjects were children in the fourth, fifth, and sixth grades. Each group replication involved only children from a single grade.

Not all of the subjects included in the study reported in Chapter 3 were included in the present study. The reason for this is that the procedures for recording attention data had to be modified as experience in the early part of the data-collection procedure demonstrated inadequacies in them. The data available thus shows differences in the number of cases involved from one learning conditions to another and from one attention-measuring procedure to another.

Results

The extent to which attention measures agree is provided in Table 4.01. When the four learning conditions were combined, the correlation was .50 (N=153). When one considers the fact that the two measures of attention were produced by independent observers, the reliability of the measures is surprisingly good.

Evidence is provided in Figure 4.01 to show that visual orientation to the task relevant stimuli declined with the passing of time. It is obvious that the decrease was greater within days than it was between. Of particular interest is the fact that the gradient representing declining visual "attention" for Monday was not as steep, nor does it dip so low as on Tuesday and Wednesday.

Table 4.01

BISERIAL CORRELATIONS BETWEEN AN OBJECTIVE MEASURE OF ATTENTION AND AN IMPRESSIONISTIC MEASURE OF ATTENTION

Learning Condition		N	r	p
I	________________ "No, that's wrong"	38	.74	<.01
II	"That's right" "That's wrong"	39	.39	<.01
III	"That's right" "No, that's wrong, ____means___."	30	.45	<.01
IV	________________ "No, that's wrong, ____means___."	46	.38	<.01

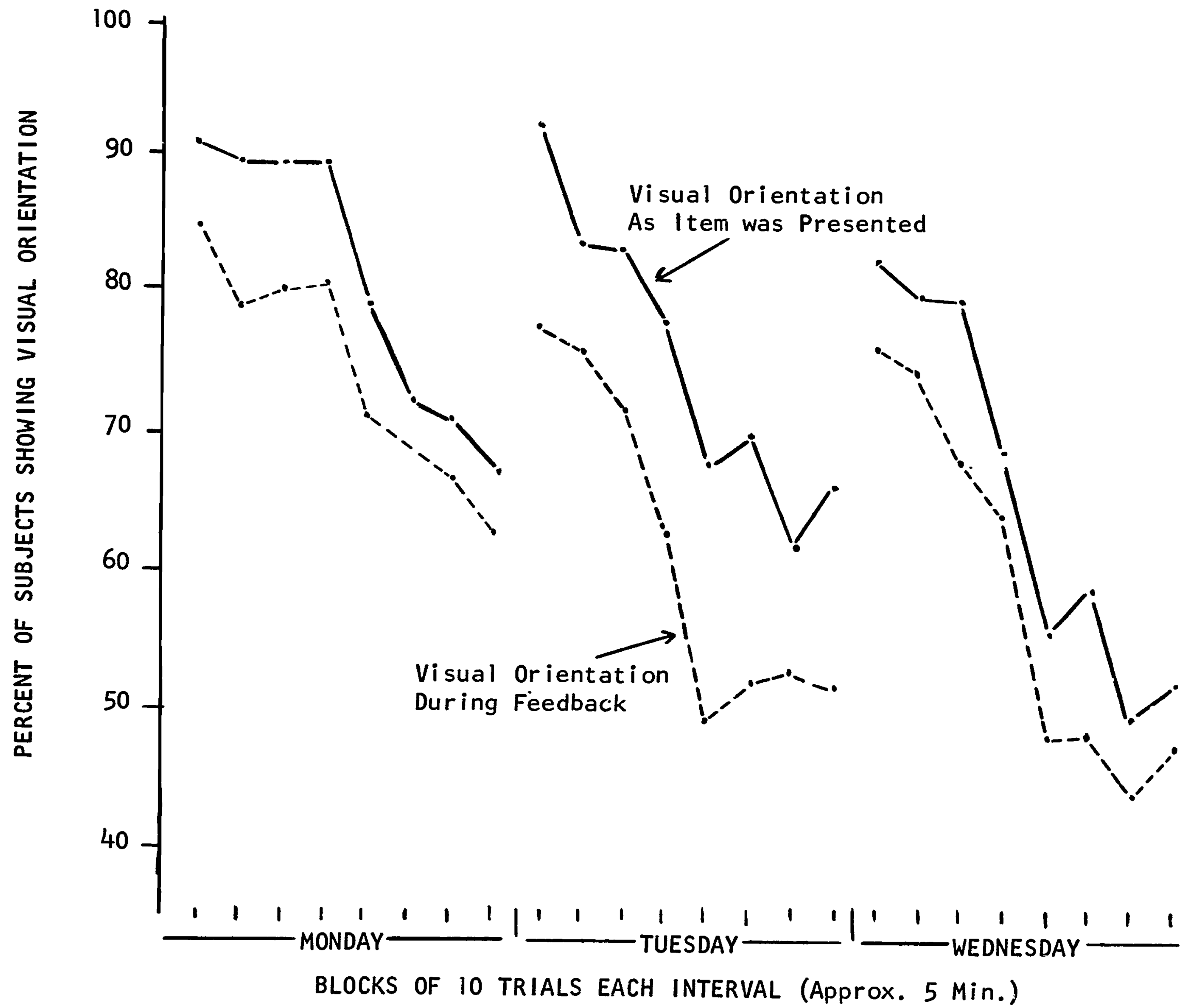


Figure 4.01. Objectively Measured Visual Attention Where 220 Separate Measures were Taken for Each Interval (Block of 10 Trials x 22 Groups).

The visual attention decrement occurring within a single day cannot be directly shown with the present data to affect retention of the task. This is due to the fact that the twenty items, constituting one day's learning task, were repeated four times through the course of the trials. A decrement in attention could not be reflected in a reduction of measured learning since a presentation near the end, say at the 70th or at the 80th trial was a word already encountered three times previously in the acquisition series for that day.

Relationship of "Attention" to Learning. Individuals learning scores separated according to feedback condition, were correlated with the objectively measured attention values. For this purpose the two objective measures of attention were combined. The correlations are given in Table 4.02. Of interest to note is the fact that the objective measure of attention was correlated positively with achievement only in the case of the first two learning conditions which required that the subjects obtain information via the visual channel if learning was to occur with maximum effectiveness. The negative correlation in the case of Condition 3 is not readily understood.

In Table 4.03 the same relationships shown in Table 4.02 are given but with the subjects classified separately into those who responded part of the time to the experimenter's questions and those subjects who did not. These correlations lack sufficient stability for any pattern to be evident.

A second, different type of attention measure had been obtained which depended upon the experimenter's general impression of who was attending and who was not. Correlations were computed between the amount of learning

Table 4.02

CORRELATIONS BETWEEN OBJECTIVELY MEASURED ATTENTION AND LEARNING UNDER EACH OF FOUR KNOWLEDGE-OF-RESULTS CONDITIONS

Learning Condition		N	r	t	p
I	____________________	38	.28	1.75	.05
	"No, that's wrong"				
II	"That's right"	39	.21	1.31	.10
	"That's wrong"				
III	"That's right"	37	-.22	1.38	.10
	"No, that's wrong, ___means___."				
IV	____________________	46	-.03	0.20	--
	"No, that's wrong, ___means___."				

Table 4.03

CORRELATIONS BETWEEN OBJECTIVELY MEASURED ATTENTION AND LEARNING UNDER EACH KNOWLEDGE-OF-RESULTS CONDITIONS, SUBJECTS DIVIDED ACCORDING TO WHETHER SUBJECTS WERE OR WERE NOT RESPONDING PART OF THE TIME

		Subjects Responding 1/4 the time	Subjects Not Responding at any time
I	________________	.67** (N=20)	.28 (N=18)
	"No, that's wrong"		
II	"That's right"	.09 (N=19)	.37* (N=20)
	"That's wrong"		
III	"That's right"	-.12 (N=19)	-.29 (N=18)
	"No, that's wrong, ___means___."		
IV	________________	.20 (N=24)	-.08 (N=22)
	"No, that's wrong, ___means___."		

* Significant at .05 level
** Significant at .01 level

and the experimenter's impressionistic judgement of which Ss are shown in Table 4.04.

These ratings were produced by E at the conclusion of each learning session and without communication with the observing experimenter who made the objective recordings of visual attention. Again, as with the objective measures, the auditory feedback conditions which provided larger amounts of information show no correlation of attention with amount learned. Where so little auditory information was given that visual attention was essential to learning, there did exist substantial correlations (conditions 1 and 2).

Discussion

While the study presented in this chapter is peripheral to the main sequence of studies, it is reported because it deals with a problem of central importance to classroom management. Teachers want pupils to "attend", particularly during recitation sessions and the common criterion of attention is whether the gaze of the pupil is directed towards the teacher or towards the discussant or towards whatever is the central source of information. The data reported indicate that this criterion of attention may have some limited utility when the source of information requires the intake of visual information. However, when the source of information is auditory, or the transmission of visual information is redundant or unnecessary, then this criterion of attention has no value. The emphasis which has been placed on the position of the eyes in determining whether a child is or is not attending appears to be unrealistic.

Rather than being preoccupied with the direction of the eyes, the

Table 4.04

BISERIALS CORRELATIONS BETWEEN IMPRESSIONISTIC JUDGMENTS OF ATTENTION AND LEARNING UNDER EACH OF FOUR KNOWLEDGE-OF-RESULTS CONDITIONS

Learning Condition		N	r	p
I	________________	52	.59	<.01
	"No, that's wrong."			
II	"That's right"	60	.40	<.01
	"No, that's wrong."			
III	"That's right"	60	.03	
	"No, that's wrong, ___means___"			
IV	________________	68	.00	
	"No, that's wrong, ___means___"			

teacher should be concerned with three aspects of processes related to attention. First, there is the matter of whether the pupil can, under the given circumstance receive the message that is being transmitted to him. Except for the case of the transmission of visual information, the teacher can do little to determine whether the information is or is not being received, but he can make sure that there are no obvious interfering circumstances such as competing sources of information, noise, and so forth. Second, the teacher should be concerned with whether the level of arousal of the learner is sufficient to produce effective learning. The only objective criterion is whether the pupil is asleep or awake, but conditions can be arranged in the classroom so that there is sufficient activity to ensure an adequate level of arousal. Third, the teacher must be concerned with the internal condition of the pupil and whether it is such that he is utilizing and storing the information provided. This can only be determined by checking on the pupil and determining how well his information-using processes have been working.

CHAPTER V

EFFECTIVENESS OF PUPILS AS REINFORCING AGENTS

The Problem

Teachers in most elementary school classrooms have paired pupils for various activities in their daily instructional experiences. This study was an attempt to investigate this classroom procedure as a method of instruction.

This was a study of verbal learning and reinforcement in a simulated classroom situation with four different learning conditions involving the interaction of pupils one with another and the interaction of individual pupils with the task alone.

The four conditions of learning established for this study were described as follows: Conditions I, II and III all involved pairs of pupils working together on the task. Condition IV involved the pupil working with the task alone.

Condition I and II occurred in the same set of pupil pairs. One member of the pair (Condition I) assumed the teacher's role and retained that role throughout the learning task. The other member of the pair (Condition II) assumed the pupil role also throughout the learning task.

Condition III also involved a treatment where pairs of pupils worked together. One pupil assumed the pupil role and one pupil the teacher role; these roles were reversed at the mid-point of each task.

Condition IV was an isolation type of treatment in which the subjects did not interact verbally in pairs as in all other conditions. In this condition the subject worked with the task as a self-instruction device which the subject used and controlled by himself.

The interaction of pairs using the task and the interaction of pupils with the task alone were completed without reinforcement procedure

by the experimenter. Reinforcement came from the pupils within the pairs and from the task.

General Statement of Purpose

The intent of the present study was to answer certain questions regarding the effectiveness of learning under conditions of individualized instruction in which pupils teach pupils and in which pupils work alone.

The study was undertaken in a setting which simulated a conventional classroom. The teacher-experimenter was not given a role of teaching, but served instead as an observer, recorder, and as a resource person for questions about procedure. Pupils assumed the role of teacher or pupil and worked directly with each other in a one-to-one relationship. The classroom became a complex of pupil-pupil interaction.

In addition to the investigation of the methodology above, this study was a follow-up of studies involving teacher-pupil interaction, both direct and vicarious, described in previous chapters. The study used the same task. Comparative data will be presented in a later chapter.

This study attempted to answer the following:

1. Does reinforcement of correct responses to a given rote memory task by a fellow pupil produce learning comparable to that produced by teacher reinforcement? In what measurable amounts does this learning occur? and how well is it retained over a brief period of recall?

2. What learning takes place in pupils acting as the teacher-reinforcer? Does the fact that they give the reinforcement rather than receive it make a difference in learning this task?

3. If the roles of a pupil-pair, those of pupil and teacher, are reversed for one-half the task, will the resulting change in role produce better rote learning?

4. What relationships produces the most effective learning for both members of the interacting pairs?

5. What learning results when the pupil reacts only to the rote learning task without interacting with another person? (As in the situation which takes place with a teaching machine.)

6. Is the social interaction reinforcement mode more effective than reinforcement provided directly by the task materials?

7. How much time is required to complete the learning task in each condition? Is one mode of learning more efficient than the others?

8. Does age (grade level) affect this kind of learning? If so, in what ways?

Procedures

The establishment of procedures used in selecting and identifying subjects, a description of the task, procedures used by the experimenter in each learning condition, and methods of collecting data are presented in this chapter.

Subjects. Subjects were fourth, fifth and sixth graders in the public schools of Salt Lake City. A total of 208 subjects was utilized with data from 192 included in the final report of the study. The

additional two groups of eight subjects were used to supplement groups where absenteeism did not allow one subject to complete all phases of the learning experience.

Task difficulty did not allow experimentation at a much lower grade level. It was experimentally important that the subject be naive in the task he was to learn, but capable of learning the materials involved.

A reading test score was used to equate data groups for each condition in the learning experience. A total division into twenty-four sets of eight subjects each was made before the study began (see Table 5.01). Groups were equated, as nearly as was possible, by drawing subjects with scores from above and below the median reading score of that group and placing them in equal numbers in each condition to be studied. Pupils with German language background or previous German instruction were not used in the study.

The Task to be Learned

The learning tasks consisted of sixty German stimulus words which were to be correctly associated with one of two English responses. The correct response appeared on one side of a card which had the stimulus word plus the two possible responses on the opposite side. The task is the same as that described in the introduction to this report but it involved the following reinforcing conditions:

Condition I & II. These conditions of learning involved the pairs of subjects whose roles remained constant in the study. Condition I was the pupil acting as the teacher who presented a card showing the

TABLE 5.01

READING SCORE PLACEMENT VALUES UPON WHICH PUPIL-SETS WERE ESTABLISHED

		Grade Four		Grade Five		Grade Six	
		SETS					
		1	2	3	4	5	6
		METROPOLITAN		READING	SECTION	TEST	
Condition I	1	4.9	3.9	5.7	3.5	7.9	6.1
	2	7.2	6.1	5.6	10.0	2.7	6.1
	3	5.3	3.6	5.1	7.3	3.2	6.8
	4	5.1	6.3	4.9	3.5	2.7	7.2
	5	4.2	2.5	5.9	7.3	7.7	3.9
	6	4.4	3.3	6.3	4.8	6.8	4.0
	7	4.9	7.2	6.6	7.7	3.6	4.4
	8	7.9	7.2	6.6	4.2	3.9	4.9
	Total	43.9	40.1	46.7	48.3	38.5	43.4
Condition II	1	3.9	3.8	5.3	4.4	7.9	6.1
	2	5.7	6.8	4.9	9.2	3.9	6.3
	3	5.3	3.3	5.1	10.0	3.2	3.9
	4	7.9	6.3	4.9	3.0	2.3	7.2
	5	4.2	4.0	6.1	7.7	6.3	4.4
	6	4.2	4.0	6.1	4.5	6.3	4.2
	7	4.5	6.8	6.6	9.2	3.8	4.5
	8	6.3	7.9	6.6	3.7	3.4	5.1
	Total	43.0	42.9	45.6	51.7	37.1	41.7
Condition III	1	3.9	4.4	6.8	5.3	5.7	5.1
	2	4.0	5.1	7.1	5.1	5.5	5.7
	3	6.3	4.4	4.9	5.1	5.7	6.1
	4	6.1	4.7	4.5	4.7	5.5	7.9
	5	3.9	4.7	7.3	5.3	4.9	6.3
	6	4.0	5.3	7.7	4.9	4.3	7.2
	7	6.1	5.5	4.7	4.9	4.7	7.7
	8	7.2	6.3	4.7	4.5	4.2	7.9
	Total	41.5	40.4	47.7	39.8	40.5	53.9
Condition IV	1	3.7	4.0	8.0	5.7	3.8	7.9
	2	3.8	4.0	8.0	5.5	4.5	7.9
	3	4.9	5.1	8.4	5.5	4.9	3.2
	4	4.7	4.0	8.4	6.1	5.1	3.8
	5	7.2	4.0	3.8	6.3	6.3	7.9
	6	7.7	4.2	4.0	6.6	5.3	3.5
	7	7.2	4.2	4.2	6.6	5.5	7.9
	8	3.4	4.3	4.2	7.3	5.7	3.8
	Total	42.6	33.8	49.0	49.6	41.1	45.9

stimulus word to the subject acting as the pupil (Condition II). The pupil selected one of the response words and the pupil-teacher stated, "the right answer is ____________." This statement was made in all cases regardless of the correct or incorrect responses of the subject.

Condition III. The subjects followed the same procedures here as in Condition I & II but reversed roles after the daily study was one-half completed.

Condition IV. In this condition the subjects did not receive feedback from a "teacher" but they did receive feedback from the cards having the stimulus word and responses on one side and the statement "the right answer is ______" on the reverse side. These subjects worked with the cards. They sub-vocally made a response to the stimulus word and then reversed the card to read the reinforcing statement. A representation of the four conditions is found in Figure 5.01.

The task was presented on three-by-five cards. The cards used in Conditions I, II and III were identical on both sides except that on the teachers side the correct answer was marked with an asterisk (Figure 5.02).

The cards used in Condition IV were identical to those in Condition I and had the reinforcing statement on the opposite side (Figure 5.03).

Data was gathered with one testing device, a multiple choice test previously described through which the material for each day was tested. The test was administered on Friday to the groups completing the learning task that week. The same test was used to provide retention data on a re-test of subjects. This re-test was given to all subjects on the same day and gave a retention score with delays of nine, eight,

Condition 1 — Teacher role throughout

Condition 2 — Pupil role throughout

Condition 3 — Teacher and pupil roles reversed in the middle of the task

Condition 4 — Self instruction role

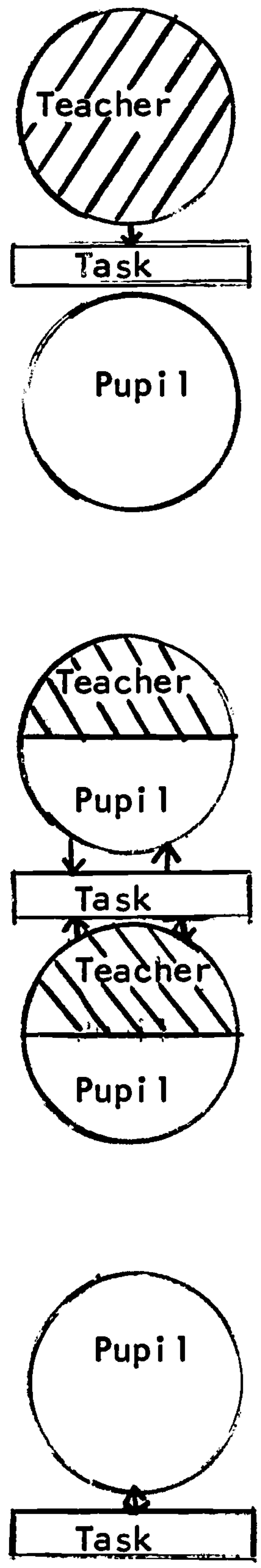


Fig. 5.01 Four conditions of learning.

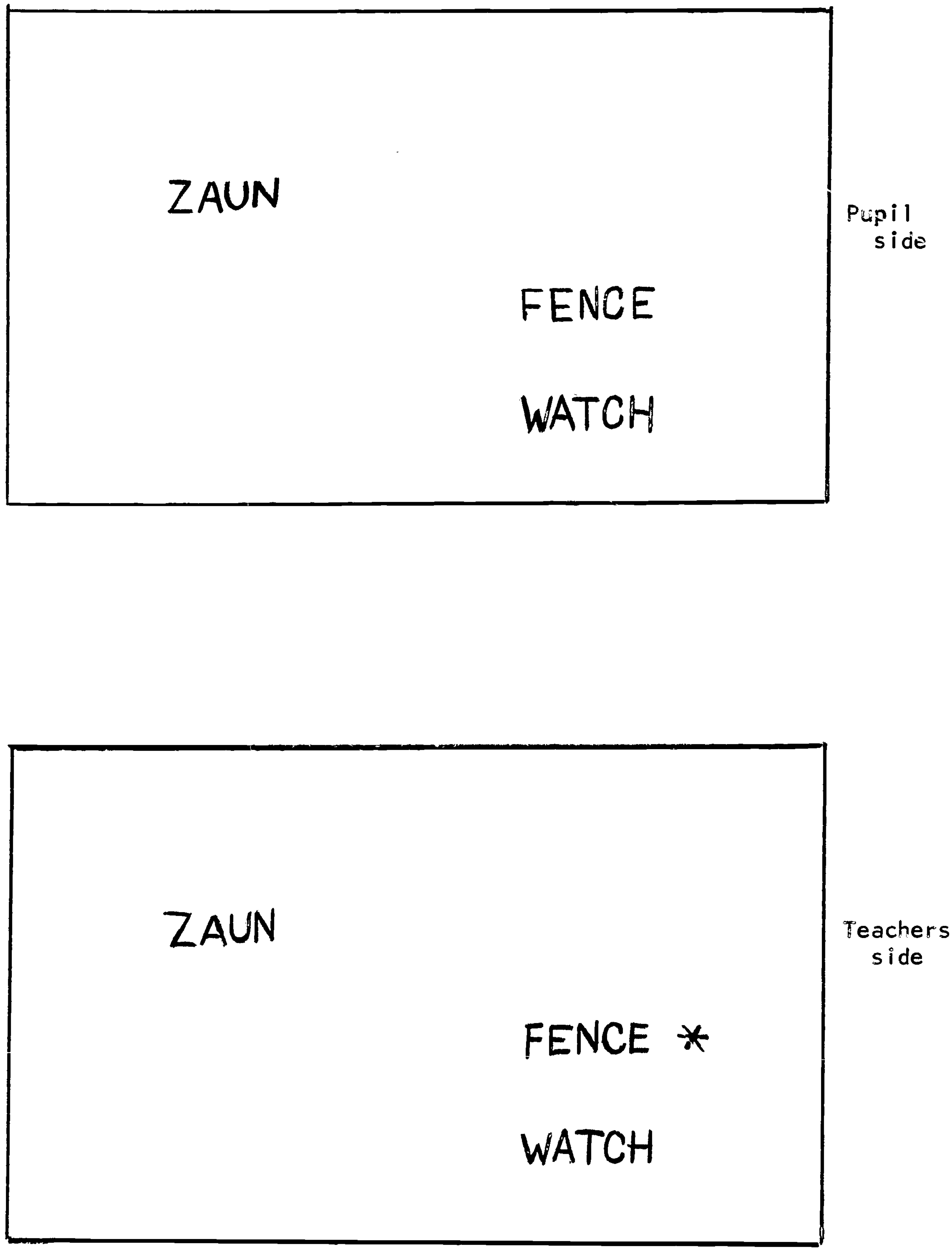


Fig. 5.02 Card used in conditions I and II.

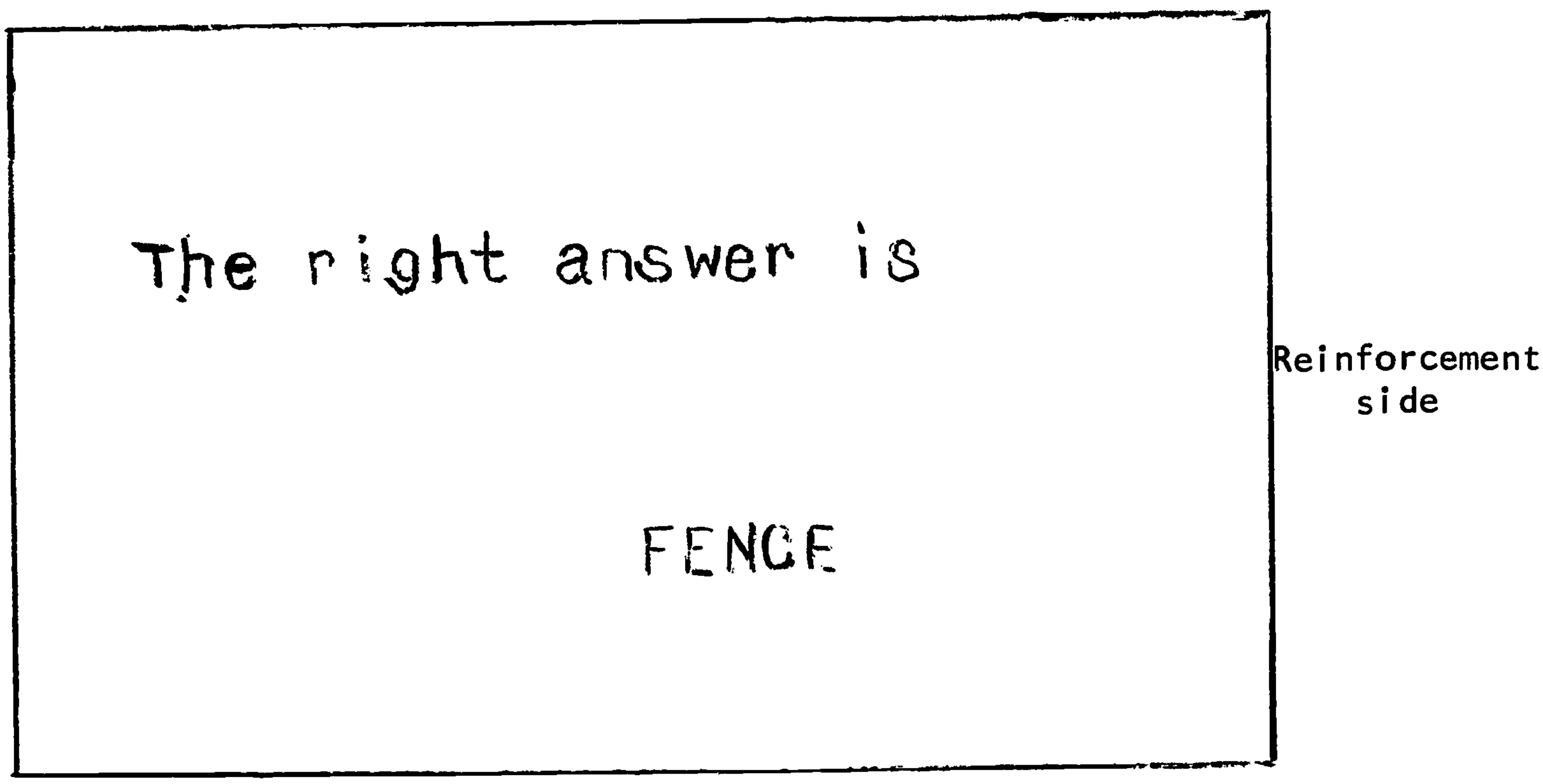


Fig. 5.03 Card used in Condition IV

seven, five, four and three weeks for eight subjects in each of the four conditions.

Only two responses were given so that immediate knowledge of correct or incorrect response was constant for all subjects.

Design

A simulated classroom situation was the setting for all experimentation. A teacher-experimenter was in the room as an observer and final authority or resource person, but was not actively engaged in teaching situations found in the four learning conditions. The teacher-experimenter gave instructions on how the experiment would be conducted and provided for all physical equipment necessary.

Subjects were paired by reading score data and four pairs met and worked together to learn the task. Subjects within the pairs were

assigned a teacher or pupil role which was maintained throughout the experiment. Other pairs reversed the pupil and teacher roles after one-half of each day's material was completed.

Eight other subjects met with the teacher-experimenter and worked individually with the task on flashcards. The reinforcement statement was printed on the back of each card.

Twenty items of the task were studied on each of three days, Monday, Tuesday and Wednesday with testing being completed on Friday.

Teacher-Experimenter Procedures

The teacher-experimenter established the classroom procedures to be used in each of the four conditions. Seating arrangements were made to suit the conditions with subjects facing one another in Condition I, II and III, and subjects seated separately in Condition IV.

General instructions to the subjects were standard for all conditions. Special instructions were also given establishing the particular condition.

Standard Instructions. The experimenter directed subjects to their proper seats and then stated, "You are here to learn some German words and what they mean in English. You will work with these cards which have the German word and two English words on them. The cards are in the order you are to use them. You will go through the cards twice. My helper (second experimenter) will work with me with these four cards and we will show you just how you are to use the cards to learn the meanings of these words. You will then go through four trial cards to be sure you understand the instructions. It is important that you watch me carefully as we do this, so that you can do it the same way.

Condition I and II. The experimenter stated, "You four children on this side of the desks will be the teachers in this group. You four on the other side are pupils. The ones serving as the teacher will show the card to the pupils. Like this (Experimenter will demonstrate by presenting the first card to his helper). The teacher will say the German word. The pupil will also say the German word and the two English words. The pupil will then choose one of the English words as the answer and say it, (experimenters demonstrate) and then the teacher will say the right answer is ______, and go to the next card. You will go through this stack of cards twice. When you finish, raise your hand. Are there any questions?" (Questions were answered) "You may now go through the four trial cards on the table. We will watch you and help you if you need help."

Condition III. The experimenter gave the same instructions as in Conditions I and II but gave instructions that the teacher and the pupil would change roles at the half-way mark before going through the stack the second time.

Condition IV. In addition to the standard instructions, the experimenter demonstrated the method used in this condition. The experimenter picked up a sample card prepared for this condition and explained to the subject, "You are to look at the card and in your mind choose one of the answers. They you will turn the card over like this and you will read to yourself, silently, the right answer is ______. You can see that you have the right answer on every card. Check your choice against this answer. Now you may go through these four trial cards."

A try-out of the design using two subjects in each of the conditions was completed to test the instructions and the design before the data for the study were collected. No significant changes in procedure were made after the try-out.

RESULTS

The data for this study was collected from four sources. The first two sources were scores from immediate and delayed retention tests of the task. The third source was a time record on each subject indicating the time taken to complete each day's task and the fourth source was the observed behavior of the subjects recorded by the experimenter in anecdotal and tabulated form. Results from these sources are reported in the following sections of this chapter: (1) the Immediate Retention Test Data; (2) the Delayed Retention Test Data; (3) the Time Record Data.

Immediate Retention Test Data

Data was gathered on each of the 192 subjects involved in this study with a test of the learning task. The test was given on Friday to subjects who had worked with the learning tasks that week.

The means and standard deviations were computed for each of the sets of subjects at each grade level in each condition and for each of the three tasks. Tables 5.02 and 5.03 were constructed to show these data. The means for Task A were higher than those of Task B and C. The means for Condition I, teacher role only, were generally lower than

TABLE 5.02

MEANS AND STANDARD DEVIATIONS

OF IMMEDIATE RETENTION TEST SCORES FOR GRADE

LEVELS, CONDITIONS, AND TASKS

Learning Condition	Grade	Task A M	Task A SD	Task B M	Task B SD	Task C M	Task C SD
I	4	8.88	3.02	8.31	3.08	7.88	2.93
	5	11.69	4.07	8.13	3.82	7.00	2.96
	6	10.19	3.17	7.19	1.98	6.81	3.03
II	4	12.31	3.16	8.63	2.67	9.88	2.87
	5	12.94	2.82	9.94	2.49	8.94	2.75
	6	12.25	2.88	9.13	2.62	9.31	2.67
III	4	13.19	3.40	9.31	3.02	8.19	3.15
	5	11.94	3.56	8.56	2.69	7.88	3.35
	6	9.88	3.33	8.13	2.50	7.94	2.77
IV	4	12.06	2.30	8.50	2.57	7.44	1.97
	5	12.75	3.91	8.19	2.27	7.81	3.26
	6	14.75	2.77	9.63	2.29	6.94	2.84

TABLE 5.03

MEANS AND STANDARD DEVIATIONS FOR CONDITIONS BY TASKS

Learning Conditions	A		B		C	
	M	SD	M	SD	M	SD
I (Teacher Role Only)	10.25	3.64	7.88	3.09	7.23	3.01
II (Pupil Role Only)	12.50	2.97	9.23	2.65	9.38	2.80
III (Teacher and Pupil)	11.67	3.69	8.67	2.78	8.00	3.10
IV (Self-Instruction)	13.19	3.28	8.77	2.46	7.40	3.12

for all other conditions. Means for each grade level did not appear to vary very much.

Means in Table 5.03 show that there are differences between conditions and even greater differences across Tasks A, B and C in each condition.

In order to examine these differences the analysis of variance shown in Table 5.04 was computed with three main effects: conditions, tasks (days), and grade levels.

<u>Main Effects</u>

The significant main effects of this study were Conditions ($p<.01$) and Tasks ($p<.001$). Grade level differences were not significant.

In order to examine the components of the variance of the third main effect Conditions, a comparison was made of each condition with each one of the other three.

These conditions were previously defined but for convenience to the reader, the following brief description was re-stated.

Condition I represented that treatment wherein one subject in a pair assumed the role of teacher and maintained that role throughout the experiment.

Condition II represented that treatment wherein one subject in a pair assumed the role of pupil and maintained that role throughout the experiment.

Condition III represented that treatment wherein both subjects in a pair were acting as teacher or pupil for one-half of the task and reversed their roles at the half-way point.

TABLE 5.04

ANALYSIS OF VARIANCE FOR CONDITIONS, GRADE LEVELS, AND LEARNING TASKS -- IMMEDIATE RETENTION TEST DATA

Source of Variation	Sum of Squares	df	M Square	F	p
Between Ss	3393	191	17.76		
Grades	9	2	4.50	0.27	
Conditions	279	3	93.00	5.58	.01
Grades X Conditions	106	6	17.67	1.06	
Residual Between	2999	180	16.67		
Within Ss	4138	384	10.78		
Tasks (Days)	1682	2	841.00	141.58	.001
Tasks X Grades	23	4	5.75	0.97	
Tasks X Conditions	134	6	22.33	3.76	.001
Tasks X Grades X Conditions	162	12	13.50	2.27	.01
Residual Within	2137	360	5.94		
Total	7531	575			

Condition IV represented that treatment wherein the subjects used a self-teaching device to learn the task and were not paired but worked alone.

This main effect produced by these conditions was also examined statistically with an individual degrees of freedom test which compared the immediate retention test scores for each condition. Table 5.05 gives information on these comparisons.

This additional analysis provided evidence that all other conditions were superior to Condition I. The comparison of Condition II and IV, the two variables with the highest mean scores with Condition III and I, the two variables with the lowest mean scores, showed a superiority in favor of Conditions II and IV at the .001 level of confidence.

When treated separately, learning in Condition III, teacher and pupil role, was superior to that in Condition I, teacher role only, and therefore Conditions II and IV were also significantly superior to Condition I, in that, the mean scores of these two conditions were even higher than the mean scores of Condition III.

In comparing Conditions II and IV, the two conditions with highest scores, it was found that Condition II, pupil role only, was significantly superior to Condition IV at the .05 level of confidence.

Although mean score comparison favored Condition IV, self-instruction, over Condition III, the difference did not reach a level of significance. Condition II was significantly superior to Condition III at .001 level of confidence. These two comparisons were not orthogonal, but the superiority was apparent from the total scores which showed Condition II to be superior.

TABLE 5.05

INDIVIDUAL DEGREES OF FREEDOM TEST COMPARING
IMMEDIATE RETENTION TEST SCORES OF
THE FOUR CONDITIONS

Source of Variation	df	Ms	F	p
Condition II & IV vs. Cond. III & I	1	550.13	33.02	.001
Condition II vs. Condition III	1	184.26	11.06	.001
Condition II vs Condition IV	1	73.50	4.47	.05
Condition III vs. Condition I	1	213.01	12.79	.001
Condition IV vs. Condition III	1	25.01	1.50	

Note: All computations based on 180 degrees of freedom because 180 were involved in the error term from Table 5.04

Interactions

Two interactions in the analysis of variance reached levels of significance as was shown in Table 5.02. One double interaction, Task x Condition, was significant. This double interaction was plotted graphically in Graph 5.01. This graph was derived from the mean scores making up this double interaction and were shown in Table 5.03. Graph 5.01 showed that Conditions I and III had a similar learning decrement across the three tasks, while Condition II and IV varied somewhat across the three tasks. Condition II showed a sharp decrement in learning from Task A to Task B, but showed a slight increment in learning in Task C. Condition IV, on the other hand, showed a very sharp decrement from Task A to Task B and this decrement continued to Task C at a much steeper decline than in any of the other conditions. The differences in decrement between conditions, across the three tasks appeared to be one reason for this interaction being significant. In order to present evidence of this decrement the difference scores between tasks for each condition were computed and were shown in Table 5.06.

The difference scores in Table 5.06 indicated the total decrement in correct responses from one day's task to the next. The difference scores demonstrated that the decrement from Task A to B was much greater than from Task B to C.

These difference scores were compared using an _a posteriori_ test, _Scheffe's Test for Multiple Comparisons_, (Edwards, 1962, p. 154), to determine if the differences varied significantly from one condition to any other. Table 5.07 was constructed to show the results of these tests and the resultant significant comparisons.

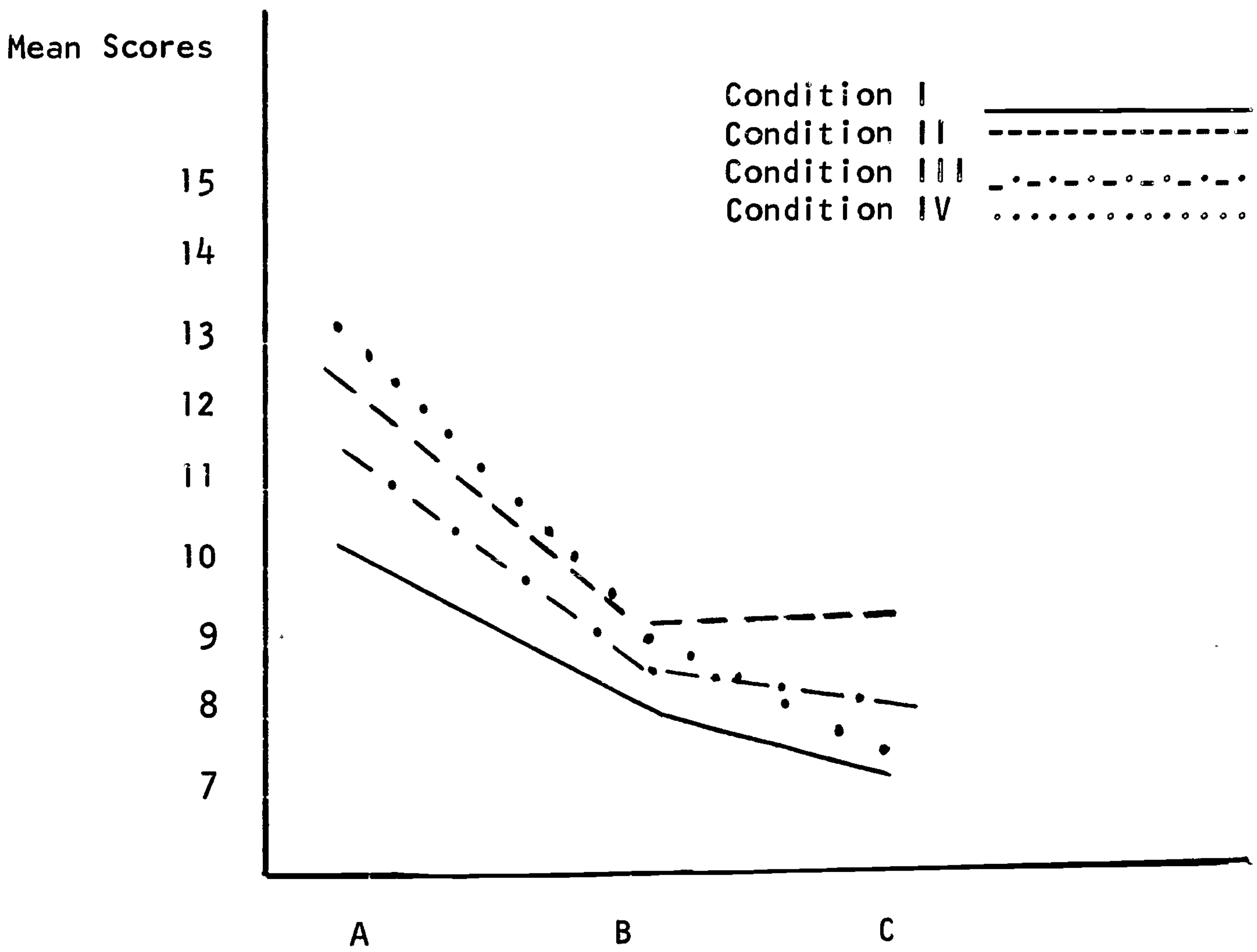


Fig. 5.04 Mean scores demonstrating double interaction of task X condition on immediate retention test data in Table 5.04.

TABLE 5.06

THE DIFFERENCE BETWEEN TOTAL SCORES OF TASKS A AND B, AND B AND C FOR EACH OF FOUR CONDITIONS OF IMMEDIATE RETENTION TEST

	Difference Between	
	A - B	B - C
Condition I (Teacher Role Only)	114	31
Condition II (Pupil Role Only)	157	-7
Condition III (Teacher and Pupil Role)	144	32
Condition IV (Self-Instruction)	212	66

TABLE 5.07

SCHEFFE'S TEST FOR MULTIPLE COMPARISONS USING DIFFERENCE SCORES FROM TABLE 5.06 TO COMPARE VARIATION BETWEEN CONDITIONS

Source of Variation	df	Ms	F	p
Condition IV vs. III from				
A to B	1	48.16	8.11	
Condition IV vs. II from				
A to B	1	31.52	5.31	
Condition IV vs. I from				
A to B	1	100.04	16.84**	.01
Condition IV vs. II from				
B to C	1	55.50	9.34*	.05

Note: All terms based on 360 degrees of freedom as in the error term in Table 5.04.
*F' needed at .05 level = 9.12
**F' needed at .01 level = 14.13

The condition which varied from the others at a level of significance was Condition IV where subjects used a self-instruction mode. The decrement in learning scores from Task A to Task B was significantly greater for Condition IV than for Condition I, teacher role only, ($p>.01$) and the decrement from Task B to C showed Condition IV with the greatest decrement of all conditions and significantly greater than Condition II, pupil role only, ($p>.05$). This final difference was also effected by the increment from Task B to C in Condition II. The difference in decrement between Condition IV and Condition II from Task A to B was not significant, and inasmuch as all other differences were less than this difference between Condition IV and II, all other differences were also not significant.

Triple Interaction

The triple interaction in Table 5.04, Tasks x Grades x Conditions was significant ($p<.01$). The source of this interaction was considered to be generally very obscure, and because most authors admit that it is next to impossible to fit this kind of interaction into any conceptual scheme, no detailed analysis of the triple interactions in this study were completed. Although some graphic or geometric presentation is suggested by Winer (1962), the complexity of a geometric three dimensional graph was not practical.

Delayed Retention Test Data

Three weeks after the final group of thirty-two subjects had completed the experiment, a delayed retention test of the tasks was given to all 192 subjects. Because of the organization of the public school

where the data were collected, it was impractical to have subjects from each grade level participate in the learning tasks each week; therefore, all subjects learning the task in a given week were from the same grade level. Thirty-two subjects, who completed the task each week, were divided equally among the four conditions of the experiment. The experiment was conducted for three consecutive weeks and then there was a break of one week to allow for the state education association meeting which curtailed school and then the experiment continued for three more consecutive weeks. When the delayed retention test was given to all subjects on the same day, this provided data which showed periods of delay in measuring retention of three, four, five, seven, eight and nine weeks. The delayed retention scores were computed for subjects in each condition to allow for comparison with the Immediate Test Data.

An analysis of variance of the data from the delayed retention test was computed and is shown in Table 5.08. This analysis omitted grade level.

The variance across Tasks A, B and C was still significant at the .001 level of confidence as it had been in the Immediate Test Data. Mean scores of the total scores for each day were plotted and the decrement from Task A to B and from B to C was apparent. The means were 9.56, 7.70, and 6.29 for Tasks A, B and C, respectively.

Evidence presented in this analysis indicated that there was superior retention of the first day's task over the other two days and that Task B or the second day's task was superior to the final task. All of these differences were significant at the .001 level of confidence.

TABLE 5.08

SOURCE TABLE FOR ANALYSIS OF VARIANCE OF DELAYED RETENTION TEST SCORES FOR THE THREE TASKS AND FOUR CONDITIONS

Source of Variation	Ss	df	Ms	F	p
Between Ss	2511	191			
Conditions	96	3	32.00	2.49	.10*
Residual Between Ss	2415	188	12.85		
Within Ss	2707	384			
Tasks (Days)	1033	2	516.50	119.00	..001
Tasks X Conditions	43	6	7.17	1.65	
Residual Within Ss	1631	376	4.34		
Total	5218	575			

*Note: An F of 2.60 was needed at the .05 level.

Conditions did not vary as they had on the immediate test data. Analysis of variance indicated significance between the .10 level and the .05 level for differences among learning conditions.

In order to examine the retention curve of the subjects across the six delay periods, the mean for each week was computed. The means were 24.87, 21.60, 24.30, 24.30, 22.56 and 23.73 for weeks three, four, five, seven, eight and nine, respectively. These means have been plotted in Graph 5.02 which shows that the retention curve was basically flat. The difference between the highest and lowest means in this graph was only 3.27 with all other differences being less than 3.27.

A table of means was constructed showing the mean scores of each group of eight subjects in each condition on both the Immediate Retention Test and the Delayed Retention Test. These means were computed for subjects in each condition who were involved in each period of delay, thereby giving a comparison of the means for each set of eight subjects on the immediate and delayed tests. These means shown in Table 5.09 were used to develop a set of adjusted means from the means of the delayed retention scores. This was accomplished by finding the difference between each immediate test mean and the average of the immediate test means and using that difference to adjust the delayed means. What this does is to correct each mean on the delayed test for the degree to which the group showed high scores or low scores on the immediate retention test. These adjusted means were then averaged by pairs across the six periods of delay to compute the average adjusted means. In effect, this was a way of reducing the variability between sets and visualizing a truer picture of what the actual retention curve for each

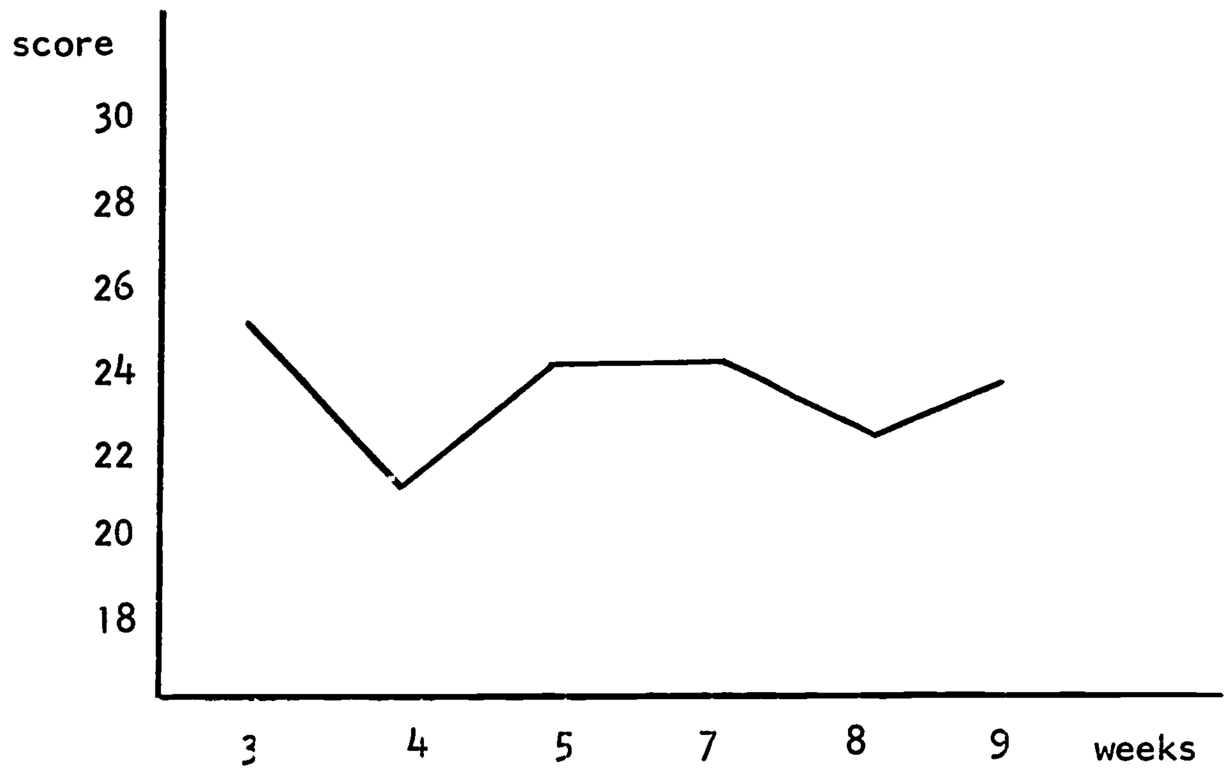


Fig. 5.05 Mean scores for delayed retention test for total scores for subjects in each of the six periods of delay.

TABLE 5.09

MEAN SCORES FOR THE IMMEDIATE AND DELAYED RETENTION TESTS, THE ADJUSTED DELAYED RETENTION MEANS, AND AVERAGE OF PAIRS OF ADJUSTED MEANS FOR EACH CONDITION ACROSS PERIODS OF DELAY

	Weeks					
	3	4	5	7	8	9
Condition I (Teacher Role)						
Immediate	25.62	22.74	24.70	29.01	25.38	24.60
Delayed	22.38	18.39	21.75	27.00	19.74	19.50
Adjusted Delayed	22.08	20.97	22.47	23.31	19.68	20.22
Average of each pair of adjusted means	21.52	21.72	22.89	21.49	19.95	
Condition II (Pupil Role)						
Immediate	32.58	28.74	30.63	33.00	31.62	30.00
Delayed	26.64	23.64	21.87	26.01	22.14	23.37
Adjusted Delayed	25.17	26.01	22.35	21.12	21.63	23.48
Average of each pair of adjusted means	25.59	24.18	21.74	21.38	23.05	
Condition III (Teacher & Pupil)						
Immediate	26.49	25.38	27.63	29.13	30.24	31.14
Delayed	23.01	20.49	24.12	25.26	24.87	28.14
Adjusted Delayed	24.87	23.46	24.84	24.48	22.98	25.35
Average of each pair of adjusted means	21.92	24.15	24.60	23.73	24.67	
Condition IV (Self-Instruction)						
Immediate	33.39	29.25	31.50	26.01	29.01	27.00
Delayed	26.88	23.88	29.49	19.26	23.49	23.88
Adjusted Delayed	22.86	24.00	27.36	22.62	28.85	26.25
Average of each pair of adjusted means	23.43	25.68	24.99	25.74	27.55	

condition would be. These delayed retention curves for each condition were then plotted and are shown in Graph 5.03 as well as the average adjusted means to show the flattened effect on the retention curve.

Although all four conditions showed a slight low in retention for subjects who completed the test after a delay of four weeks, the basic adjusted curve was fairly flat with differences between weeks of delay consistently small.

A comparison of the loss of scores from Immediate Retention Test to the Delayed Retention Test was also made by computing the total scores for each condition on each test and finding the percentage of loss for the total group and for the subjects in each condition.

Table 5.10 was constructed to show this data and this table illustrates that the greatest percentage of loss took place in Condition II, pupil role only, with a loss of 23.0 per cent from one retention test to the other. The least per cent of loss occurred in Condition III with 14.2 per cent; this condition was that one wherein the teacher-subject and pupil-subject reversed their roles one-half way through each day's learning task. The total loss in correct response scores from one test to the other for all 192 subjects was 958 which represented a percentage loss of 17.5 per cent.

Time Record Data

One of the major sources of data for this study was a time record which was made on each subject or pair of subjects as these subjects completed each of the three tasks. The time record showed in minutes the time taken to complete each of the Tasks A, B and C. Evidence of differences by condition, task, and grade level were sought. The means

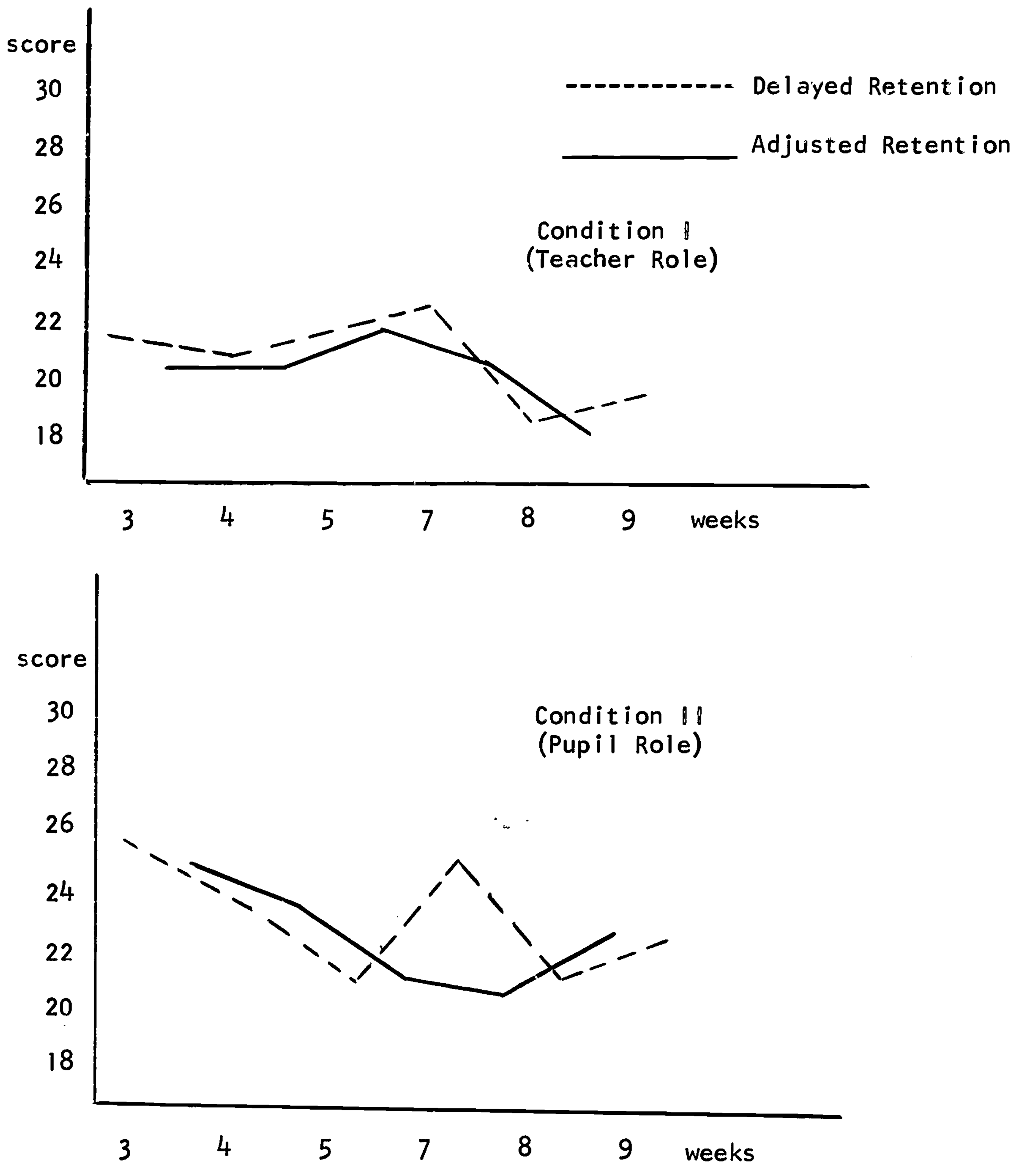
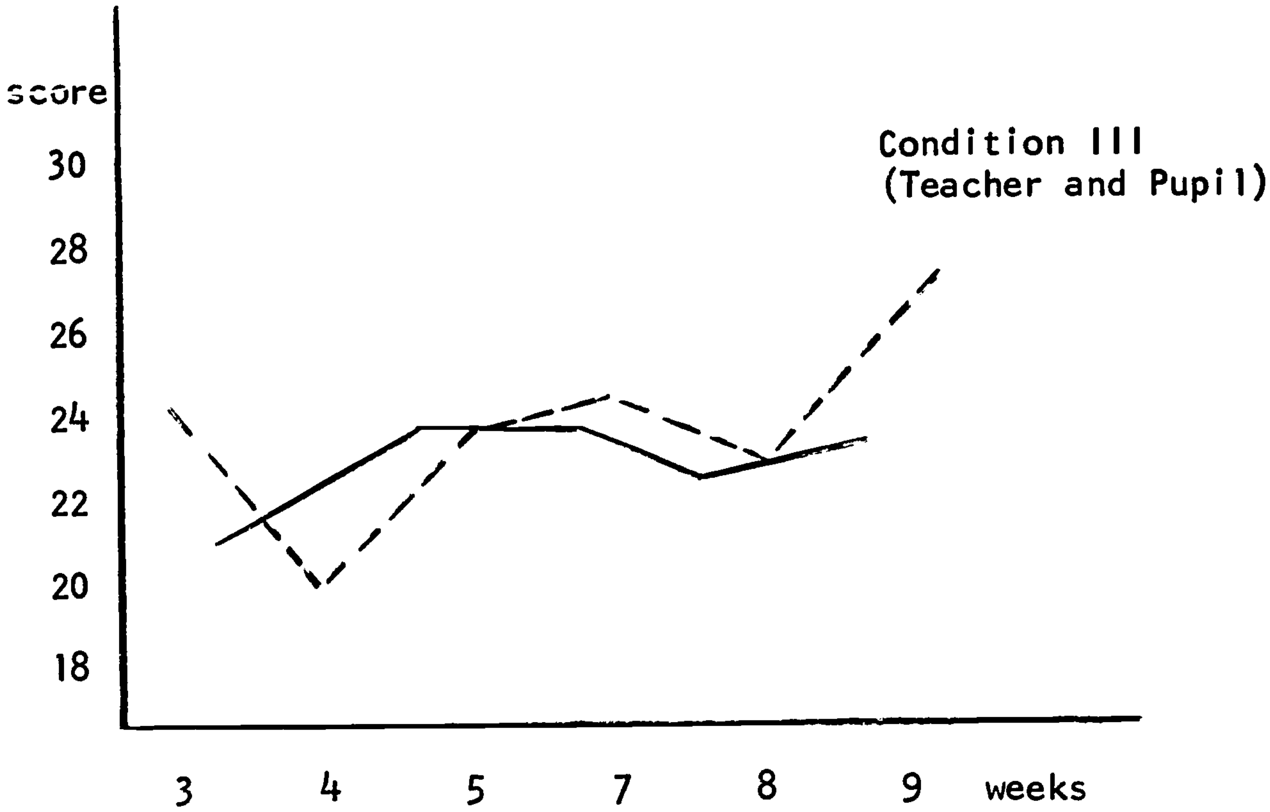


Fig. 5.06 A comparison of the delayed retention means for each of the six periods of delay and the average of pairs of adjusted means (from Table 5.09).

FIGURE 5.06 (Continued)



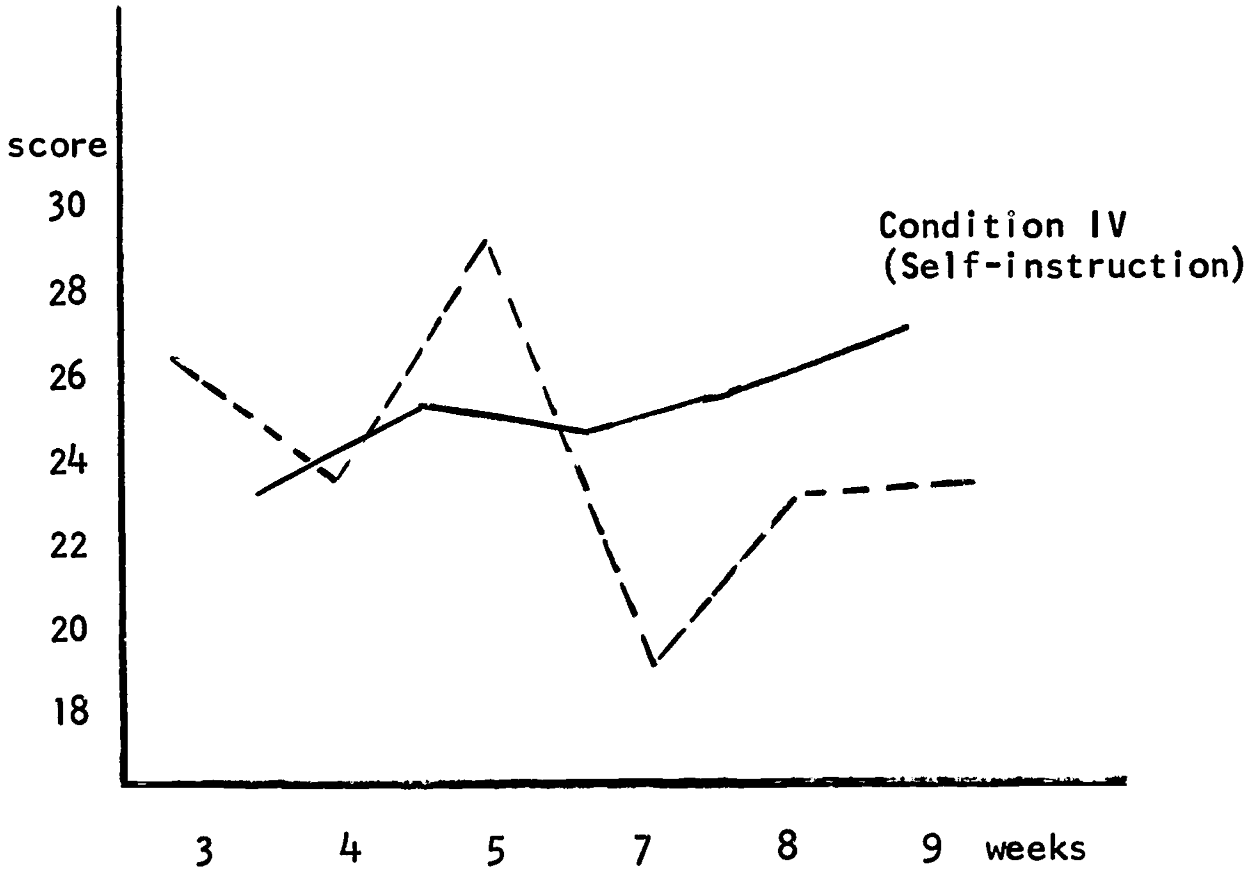


TABLE 5.10

DIFFERENCE BETWEEN TOTAL SCORES FOR IMMEDIATE AND DELAYED RETENTION TEST AND THE PERCENT OF LOSS FOR CONDITIONS I, II, III AND IV (N = 48 FOR EACH CONDITION)

	Immediate	Delayed	Difference	Percent of loss
Condition I	1217	1030	187	15.4
Condition II	1493	1149	344	23.0
Condition III	1360	1167	193	14.2
Condition IV	1409	1175	234	16.6
Total	5479	4521	958	17.5

and standard deviations for these data were computed and are shown in Tables 5.11 and 5.12.

Means and standard deviations were computed for each of the sets of subjects at each grade level, in each condition and for each of the three tasks. Table 5.11 was constructed to show the means across tasks and conditions by each grade level. The data indicated that each grade level shows a decrement of time required to complete the tasks with less time required to complete Tasks B or C than to complete Task A. It was also apparent from these mean times that fourth graders needed more time to complete the task than fifth graders; and fifth graders required more time than sixth graders to finish the tasks.

In Table 5.12 the mean times for each of the four conditions across the three tasks were computed. Conditions I and II represented subjects working together so their times were basically identical; the slight variations were due to necessary substitution of subjects because of absenteeism.

Condition III, which also represented pupil-pairs working together on the Tasks, had very similar mean times to those of Conditions I and II. Conditions IV mean times were sharply lower than those of all other conditions; in fact, less than half as much time was required by subjects in Condition IV to complete the tasks than in any of the other conditions.

An analysis of variance of the time record data was completed and the source table for this analysis is shown in Table 5.13.

TABLE 5.11

MEANS AND STANDARD DEVIATIONS OF TIME RECORD DATA FOR GRADE LEVELS, CONDITIONS, AND TASKS

Learning Condition	Grade	A M	A SD	B M	B SD	C M	C SD
I	4	16.34	3.73	15.45	3.47	13.59	3.25
	5	15.23	2.93	13.54	1.82	13.46	2.01
	6	12.85	1.05	11.02	1.41	10.98	1.93
II	4	16.36	3.72	15.39	3.50	13.41	3.31
	5	15.24	2.93	13.54	1.82	13.46	2.01
	6	12.85	1.05	11.02	1.41	10.98	1.93
III	4	17.55	3.06	15.31	2.86	13.48	1.85
	5	14.69	1.63	14.34	2.39	11.16	2.60
	6	13.26	1.45	12.46	1.57	11.23	1.32
IV	4	8.04	2.28	6.82	1.60	6.15	1.20
	5	6.97	1.67	5.68	1.43	7.60	0.98
	6	8.04	1.30	5.46	1.02	4.87	1.15

TABLE 5.12

MEANS AND STANDARD DEVIATIONS FOR TIME RECORD DATA WITHOUT REGARD TO GRADE LEVELS

Conditions	Task A		Task B		Task C	
	M	SD	M	SD	M	SD
I (Teacher Role Only)	14.81	3.15	13.34	2.99	12.68	2.72
II (Pupil Role Only)	14.82	3.14	13.31	3.03	12.62	2.74
III (Teacher and Pupil)	15.17	2.79	14.04	2.59	11.96	2.26
IV (Self-Instruction)	7.68	1.87	5.99	1.48	6.21	1.56

TABLE 5.13

ANALYSIS OF VARIANCE FOR CONDITIONS, GRADE LEVELS, AND LEARNING TASKS - SOURCE TABLE - TIME RECORD DATA

Source of Variation	Ss	df	Ms	F	p
Between Ss	8311.36	191			
Grades	731.05	2	365.53	30.93	.001
Conditions	5311.05	3	1770.35	149.78	.001
Grades by Conditions	141.75	6	23.63	2.00	
Residual Between Ss	2127.51	180	11.82		
Within Ss	1469.84	384			
Tasks (Days)	501.38	2	250.69	119.95	.001
Tasks by Grades	40.86	4	10.22	4.89	.001
Tasks by Conditions	70.39	6	11.73	5.61	.001
Tasks by Grade by Conditions	101.09	12	8.43	4.03	.001
Residual Within Ss	753.52	360	2.09		
Total	9778.60	575			

Main Effects

All the main effects of the time data were significant -- grade level ($p<.001$), tasks ($p<.001$), and conditions ($p<.001$) (Table 5.13).

The time means for grade levels were 39.47, 36.23, and 31.25 for the fourth, fifth and sixth grades, respectively. The latter variance was further tested statistically by an individual degrees of freedom test to determine which grade levels were varying significantly. The results of this degrees of freedom test are shown in Table 5.14.

The results shown in Table 5.14 demonstrated that the sixth grade subjects took significantly less time to complete the tasks in this study than did the fourth and fifth grade subjects. The fifth grade subjects also took significantly less time to complete the tasks than did the fourth grade subjects. The difference in time taken to complete the tasks varied significantly from grade level to grade level with the fourth grade subjects taking the longest time, the fifth grade subjects requiring significantly less time than the fourth grade subjects and the sixth grade subjects taking significantly less time than the fifth grade subjects to complete the tasks.

The second significant main effect in the analysis of variance was the difference between time taken to complete each of the three tasks. The mean times for subjects to complete Tasks A, B and C were 13.12, 11.67 and 10.86, respectively. Task A completed on Monday required the longest time to complete. The Tasks B and C which were completed on Tuesday and Wednesday each required subjects successively less time to complete.

TABLE 5.14

INDIVIDUAL DEGREES OF FREEDOM TEST

COMPARING TIME DATA OF THE

THREE GRADE LEVELS

Source of Variation	df	Ms	F	p
Grade 4 vs. the average of Grades 5 and 6	1	350.56	29.66	.001
Grade 5 vs. 6	1	790.53	66.88	.001

Note: All terms based on 180 degrees of freedom since 180 df were involved in the error term in Table 5.13.

The final main effect, Conditions, also showed a significant variance in the analysis of variance (Table 5.13). However, in this main effect three conditions had mean times which varied only slightly while the fourth condition had an unusual variation from the other three. Conditions I and II, of course, involved the same pairs and the time score means in these two conditions would have been identical except for a substitute subject being used to complete the data in Condition I because of an absentee. Condition III, which also was a condition involving pairs of subjects, varied only slightly from Conditions I and II. The mean times for the four conditions, I, II, III and IV were 40.82, 40.75, 41.16 and 19.88, respectively.

The lack of variation between the time needed to complete the task for the conditions involving pupil-pairs was evident also in an individual degrees of freedom test comparing the time data of the four conditions. No significant differences were found between Conditions I, II and III, but Condition IV varied significantly from all other conditions.

By testing the variance between Conditions I and III, it was demonstrated that no significant difference was required in the time to complete the tasks when pupil-pairs worked on the task together. Conditions I, II and III, the paired conditions, did not vary significantly, although subjects in Condition III required slightly more time to complete the task. However, Condition IV, self-instruction, varied very significantly from the other conditions in that subjects in Condition IV required less than half the time to complete the tasks than those subjects completing the tasks in Conditions I, II and III. The level of significance for this difference was .001. Inasmuch as Condition II,

pupil role only, had the lowest mean time of Conditions I, II and III, it was only necessary to compare this condition with Condition IV, self-instruction, in order to establish the significance of variation of Condition IV to all other conditions.

Interactions

The interactions which had Tasks as one of the interaction components were all significant in the analysis of variance in Table 5.13. The double interactions Tasks x Grades and Tasks x Conditions were examined graphically to show how the components of these interactions might have varied from one another enough to become significant. A graphic presentation was used because the variation was discernible in this type of presentation. If the interaction were zero, the components in a geometric or graphic presentation would be parallel even though of different strength numerically. The variation created by the interaction would force the graphic presentation away from this parallel graphic presentation (Winer, 1962).

First, the interaction Tasks x Grades was plotted using mean scores from Table 5.11 as a basis for the graph. Graph 5.04 was constructed to illustrate this interaction. The time decrement for the fourth and fifth grade subjects from Task A to B followed a parallel or nearly parallel form and subjects from grades five and six followed a parallel time decrement from Task B to Task C. Two lines in the graph are not in parallel form: (1) Sixth grade subjects showed a sharper decrement in time required to complete the task from Task A to B; (2) Fourth grade subjects did not show a leveling off of time from Task B to C as

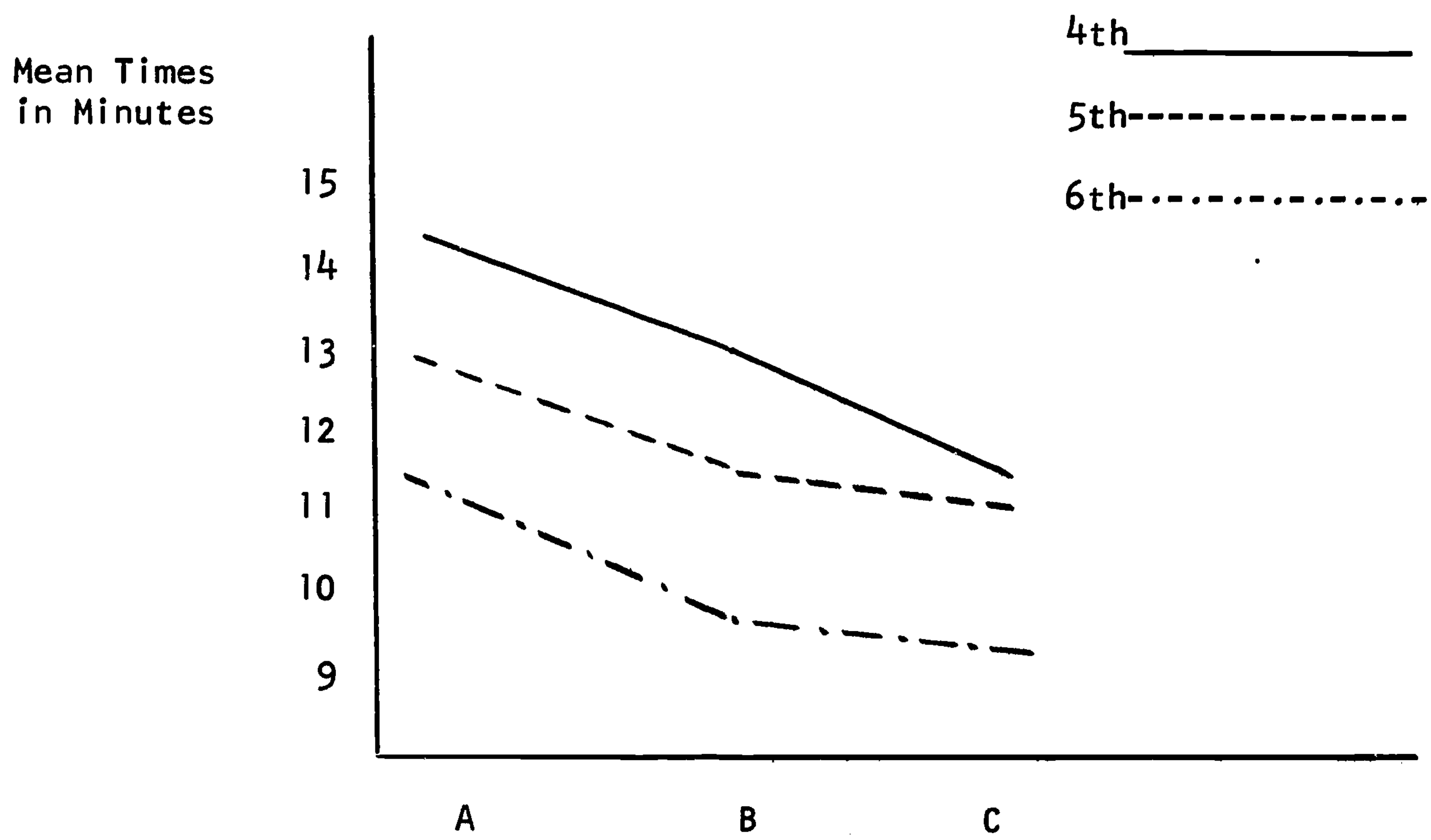


Fig. 5.07 Mean times demonstrating the double interaction tasks X grades from Table 5.11.

did subjects in grades five and six. The decrement continued down at a sharper angle for fourth grade subjects.

The second double interaction from the analysis of Time Record Data which was significant was Tasks x Conditions. This interaction was also plotted graphically using means from Table 5.12. Graph 5.05 was constructed to show this interaction. It is difficult from this graph to make any concise statement about the nature of the interaction.

Efficiency of Learning

One of the basic reasons for recording time data was to attempt to develop an efficiency of learning score. This was done for each condition and each grade level by developing the formula $ES = \frac{TT}{CR}$. (ES = Efficiency Score, TT = Total Time, CR = Correct Response on Immediate Retention Test). The fewer minutes required to learn a correct response the more efficient the learning was considered to be.

First, the formula was applied to each of the four conditions and the results were recorded in Table 5.15.

Condition IV, self-instruction, was the most efficient learning condition with a score of 0.68. All other conditions required about double the time per correct response.

An efficiency score for subjects in each grade level was also computed and recorded in Table 5.16.

Although the differences in efficiency scores by grade level were slight, a trend of efficiency related to age was apparent. Sixth grade subjects were most efficient in learning the task (ES = 1.11), followed by fifth graders (ES = 1.25) and the least efficient were fourth graders (ES = 1.38).

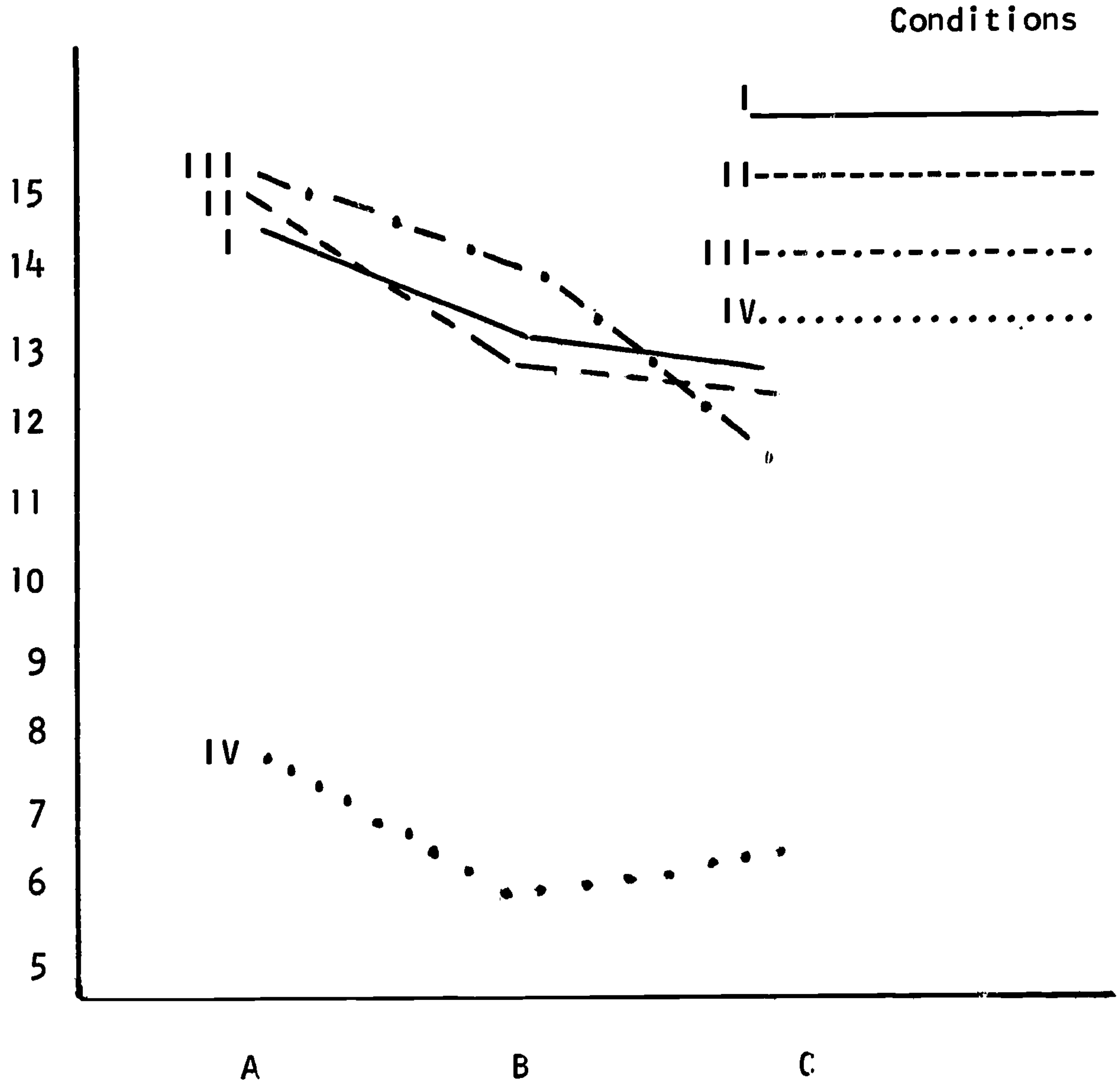


Fig. 5.08 Mean times demonstrating the double interaction tasks X conditions from time record data in Table 5.12.

TABLE 5.15

LEARNING EFFICIENCY SCORES FOR SURJECTS IN

CONDITIONS I, II, III AND IV

Condition	Total Time	Correct Responses	Learning Time Per Correct Response
I (Teacher Role Only)	1959.3	1217	1.61
II (Pupil Role Only)	1956.0	1493	1.31
III (Teacher & Pupil Role)	1975.7	1360	1.45
IV (Self-Instruction)	954.0	1409	0.68

TABLE 5.16

LEARNING EFFICIENCY SCORES FOR SUBJECTS

IN GRADES FOUR, FIVE AND SIX

Grade	Total Time	Correct Responses	Learning Time Per Correct Response
4	2526.3	1833	1.38
5	2318.4	1854	1.25
6	2000.3	1794	1.11

EXPERIMENTER OBSERVATION OF THE SUBJECT BEHAVIOR DURING LEARNING OF THE TASK

In addition to the data collected by formal instruments such as test scores and timing scores, the experimenter made a written record of the behavior of subjects which varied from the necessary procedural behavior required to learn the task.

During the first two weeks of the study, the experimenter made extensive notes of the behavior beyond the expected procedural behavior. At the end of that time, the experimenter classified the notes taken and gave categorical titles to the behaviors that appeared two or more times in his notes. The behaviors were defined by the verbal or motor activities which were observed.

In Conditions I, II and III, where pupils were paired, much more overt behavior was noted than in Condition IV where pupils worked on the task alone. The only notable factor recorded concerning Condition IV was the factor of competition between subjects on speed in completing the task first, and this occurred with only six of the forty-eight subjects in that condition.

Table 5.17 was constructed to show these recurring behavioral incidents and the number of times these incidents were observed and noted. Because of the number of subjects involved, no breakdown was made of behavior by condition, but because all subjects in a given set were from the same grade level a breakdown by grade level was possible.

The results of these tabulated forms of behavior were analyzed to give some indication of the types of behavior which were most prevalent at a given grade level.

TABLE 5.17

BEHAVIOR OTHER THAN EXPECTED PROCEDURAL BEHAVIOR OBSERVED

AND RECORDED BY THE EXPERIMENTER

Item No.	Observed Behavior	Grade Level 4th	5th	6th	Total per item	% of Grand Total
1.	Competition - based on time to complete the task (speeding-up of pace. Comments such as "Hurry up! We're ahead." etc.)	44	35	17	96	16.1
2.	Verbal interactions other than the task within or between pairs. (I miss that everytime." "Okay" "Don't talk so loud." "I thought that was it." "How many more?" "That was easy." etc).	25	27	23	75	12.6
3.	Questions asked by subjects about procedure before beginning the task.	43	19	12	74	12.4
4.	Distraction peer pressure. (Looking at peer's task to see how they compared in amount completed)	15	27	22	64	10.7
5.	Anxiety over incorrect responses. (Shifting position, frowning, striking table with hands, tapping with feet or hands, groaning)	16	19	22	57	9.5
6.	Laughter, comment or anxiety about pronunciation.	11	12	20	43	7.2
7.	Pupil-subject correction of teacher mistake (Pronunciation, procedure, checking, e.g. "Did you say baum?").	25	6	11	42	7.0
8.	Subject concern with experimenter observation. (Looking directly at the experimenter to see if they are being observed).	10	16	12	38	6.4

TABLE 5.17 (Cont.)

Item No.	Observed Behavior	4th	5th	6th	Total per item	% of Grand Total
9.	Subject-teacher making corrections of pupil work or giving directions to pupil.	10	11	13	34	5.7
10.	Pupil-subject giving response without oral verbal cue from the teacher.	21	4	4	29	4.9
11.	Obvious and overt lack of attention to the task-card at the time of the reinforcement statement by the teacher. (Looking away from the task completely at the time of feedback.)	15	5	1	21	3.5
12.	Pupil-subjects (Cont.II) who want to have a turn at being teacher and ask experimenter if they could be the teacher.	5	0	2	7	1.2
13.	Subjects asking to go through the task an additional time even though instructed explicitly on the number of times to do the task.	4	1	2	7	1.2
14.	Subject asking for experimenter's help during learning task.	3	2	0	5	.8
15.	Subject serving as teacher in a pair not taking charge and getting procedure started. (Requiring prompting from experimenter.)	3	2	0	5	.8
	Grand Totals	250	186	161	597	100%

Generally, the results indicated much more variant behavior at the fourth grade level than at the other grade levels. A total of 250 incidents were recorded at the fourth grade level, and while this was a small number compared to the procedural interactions which took place between all subjects at a given grade level (15,360), the number for fourth graders was sixty-four more incidents than those observed in the same numbers of fifth grade subjects and eighty-nine more incidents than those observed in sixth grade subjects. The fourth grade subjects had 41.8 percent of the total incidents recorded for the three grades.

The results also showed patterns of behavior related to grade level. Fourth graders generally were interested, not only in following procedure, but also in following the progress of their fellow fourth-graders during the task. Fourth grade pupils were observed showing more incidents of behavior, other than expected behavior, in the areas of (1) competition; (2) asking questions about procedure; (3) pupil correcting the teacher's errors; (4) giving a response without waiting for verbal cue; and (5) inattention at the time of reinforcement. Fourth grade subjects were also higher in incidents of observed variant behavior than fifth and sixth grade subjects in items 12, 13, 14, and 15 of Table 5.17, but the number of incidents form a very small percent of the total number of incidents. Fourth grade subjects observed behavior generally followed a pattern of non-conformity with several types of variant behavior recorded.

Fifth grade subjects recorded more incidents of behavior, other than procedural, than the fourth or sixth grade in the areas of, (1)

verbal interactions other than the task, (2) distraction by peers, and (3) concern with experimenter observation, but in all three cases these were only slightly higher. Fifth graders' observed behavior generally established a pattern of concern over what others in the room were doing.

Sixth grade subjects showed higher incidents of overt behavior in the areas of, (1) laughter or anxiety about pronunciation, (2) anxiety over incorrect responses, and (3) teacher correction of pupil work. These were also only slightly higher than those recorded for the other two grades. A pattern of conformity to expectation and anxiety about deviation from the expected was observed in sixth grade subjects.

Sixth grade subjects indicated little if any secondary need for procedural training. Results of recorded behavior in items 3, 11, 12, 13, 14 and 15 which all dealt with behavior related indirectly to procedure showed very low totals for sixth grade subjects. Sixth graders followed procedures with little error.

Total results per item of observed behavior show that subjects in this experiment were producing behavior other than that expected most often in the three following areas:

(1) Competition (item 1) ninety-six incidents.

(2) Verbal interactions other than the task (item 2) seventy-five incidents.

(3) Questions about procedure (item 3) seventy-four incidents.

Table 5.17 was organized to show the total number of incidents per item in rank order, and the three items above accounted for forty-one per cent of the recorded incidents of overt behavior for all three grade levels.

General Observations

The experimenter recorded many observations of a general nature which could not be classified in a specific category. These observations were listed here to furnish a record of these statements and to provide data of all observations, not only those which were adaptable to specific classification.

The following were the general observations, listed chronologically as these statements appeared in the experimenter's records.

1. The volume levels of the voices of each group were adjusted to fit the noise level of that group.

2. In Condition III, one pair realizing that the cards would go to the other subject after going through the stack, worked out a system wherein they placed the cards in his hands so the cards would be ready.

3. Students attended to task much more readily on second and third days. Procedure went better--less talk--less reluctance to begin.

4. More uniformity of task behavior within Condition V than any other. Less behavioral variation between Monday and Tuesday for Condition IV, self-instruction, than for any other.

5. Teacher-subject smiling at correct response seemed to serve as a reinforcer.

6. Procedures were much more fluent during second day learning trials. Pace remained more constant.

7. Anticipation of the second day's task indicated by pleasure with which student accepts the new set of cards.

8. Procedural rapport with various groups seems to affect the work attitude of the particular group of eight. The groups who asked

questions about procedures were more diligent in their study efforts during trials. Answers to questions seem to add a new "set" toward group study.

9. Third day trials were more efficient in the time and effort because of habituation of procedures. Does this affect learning? With a lessening of anxiety about procedure does learning become more effective or less effective?

10. One pair in Condition I and II followed a pattern of extreme concentrated study concentrating on each pronunciation and response. Teacher often stared at wall while making up his mind about pronunciation. Teacher's lack of action eventually caused pupil to reprimand him. Pupil wanted more rapid trials.

11. There may be some additional learning by those who finish first and attend closely to other pairs working in the room; however, this attention was very seldom evident as pupils who finished first usually talked quietly together.

15. One pair continued to keep track of errors on trials, numerically counting them.

16. Procedures seem to bother students with low reading scores. They spent much time trying to remember how they were to do it and consequently slow processes resulted.

17. Condition III, teacher and pupil, seems to present the best teacher-pupil relationship because both know they are going to be in an equal status position.

18. Very little interaction of any kind was observed in Condition IV. Some glancing at one another to see how many cards were left in each other's stack.

19. Regard for other pairs was evident in a subdued use of voice.

20. Near the end of the task some pupils became lackadaisical about their efforts in Condition I and II. Task may be boring at this point for some, or too long for slow pupils. Actions indicating boredom: not attending to card or response, slouching down in chair, looking at ceiling, avoiding task until prompted by partner.

21. Condition III seems to provide incentive for both subjects that is not evident in I and II. Interest remains high throughout task. No bored individuals or pairs observed.

22. There is no question that Condition IV, self-instruction, is an efficient procedure as time goes but there is some concern with the need for reinforcement by another person of the completed task, especially on second and third day of learning task.

23. Social prestige is probably an important factor when the subject gets to play the teacher's role.

24. Almost all sub-vocal responses in Condition IV are accompanied by lip movement and in some cases the responses are whispered.

25. Teacher-subjects at the fifth grade level appeared adept at keeping pupil on the task.

26. Enthusiasm of those in Condition III, teacher and pupil role, is remarkable. While subjects in other conditions show enthusiasm, these subjects can hardly wait to get on with the task.

27. Very few unusual occurrences were discovered at the sixth grade level. Maturity makes it possible for these people to do this task with great attention.

28. The pace was almost identical for all four pairs within each set.

of eight subjects in Condition I, teacher role only, and II, pupil role only, at sixth grade level. Practically complete attention to the task was observed throughout. Pairs finished within thirty seconds of one another.

29. The seeming maturity of sixth grade subjects in contrast to fourth and fifth grade subjects was noted.

A COMPARISON OF COMPARABLE LEARNING CONDITIONS IN THE VAN WAGENEN STUDY OF 1962 AND THIS STUDY

The task used in this study was developed and first used in 1962 in a doctoral study by R. Keith Van Wagenen at the University of Utah. His study is reported as Chapter II of this report. The same Immediate Retention Test was used in both studies.

Van Wagenen also established four learning conditions in that study which were:

Condition 1. Direct oral responding to a teacher-experimenter on every fourth item in the task. The subjects in this treatment were reinforced directly by the teacher on one-fourth of the items in the task and had a vicarious listening experience on the other three-fourths of the items.

Condition 2. Vicarious experience on all items. There were four other subjects in a set of eight, including the four in Condition 1, who never responded directly to the teacher-experimenter but only read and listened to the response made by subjects in Condition 1.

Condition 3. Was identical to Condition IV in the present study except that a teaching machine was used to present the task in the Van Wagenen study.

Condition 4. Was the same as Condition 3 above except that in addition to the feedback by the machine the teacher-experimenter also gave verbal reinforcement (no comparable condition in the present study).

Table 5.18 shows the comparisons that can be made between the data derived from the Van Wagenen study reported in Chapter II and the data from the present study.

Table 5.19 was constructed to show the means and standard deviations for each comparable learning condition with data for each study from correct responses on the Immediate Test.

The means for the comparable learning conditions were nearly identical for the subjects completing the tasks by use of flashcards or teaching machines (Comparison No. 1). Differences between the two means were very slight, but to test the significance of the difference, the means were treated with a t test. The results of this test (t = .59 with 84 degrees of freedom. t equal to 2.00 needed at the .05 level) showed no significant difference between the means of these two groups.

The second comparison which was feasible between the two studies involved subjects who at no time received verbal feedback or reinforcement. These means show in Comparison No. 2 in Table 5.19 were also tested by use of a t test and again no significant differences were noted (t = 1.73 with 84 degrees of freedom. t equal to 2.00 needed at the .05 level). The difference which neared significance in this comparison favored the vicarious subjects of the Van Wagenen study who heard the reinforcement given to others. The subjects in the present study did not hear it given to other subjects by another person, but gave the reinforcement to other subjects and thereby heard their own voices giving

TABLE 5.18

COMPARABLE REINFORCEMENT CONDITIONS FROM THE VAN WAGENEN STUDY (CHAPTER II) AND FROM THE PRESENT STUDY

Van Wagenen Study-Chapter II	Myers Study This Chapter
Condition III - Self-instruction	Condition IV - Self-instruction
Condition II - Pupil observed but did not respond	Condition I - Pupils in teacher's role observed but did not respond.
Condition I - That portion of this condition in which pupils were reinforced 100 per cent by teacher	Condition II - Pupils reinforced 100 per cent by teacher pupil.

TABLE 5.19

MEANS AND STANDARD DEVIATIONS FOR COMPARABLE CONDITIONS IN THE VAN WAGENEN STUDY AND THIS STUDY

Comparison No.	Description of Comparable Learning Condition	Van Wagenen		Myers	
		$\bar{X}$	SD	$\bar{X}$	SD
1	Subjects received task and feedback from machine or flashcard.	29.22	8.59	29.35	6.39
2	Subjects who did not receive feedback directly (vicarious or pupil as teacher).	26.30	5.87	25.35	7.17
3.	Subjects who received 100 per cent direct verbal reinforcement.	32.70	12.23*	31.10	6.34

*Only one-fourth of the items were directly reinforced in the Van Wagenen study and this mean and standard deviation represent only the scores on those items multiplied by 4.

the feedback.

Comparison No. 3 was made to show the differences between subjects who received direct verbal reinforcement from a teacher-experimenter and those who received direct verbal reinforcement from another pupil. The results of this comparison were indicated in a t test (t = 2.39). The resulting t was significant at the .05 level and favored the subjects in the Van Wagenen study who were reinforced by an adult experimenter. This comparison was approached cautiously because in this study all items were reinforced and in the Van Wagenen study the subjects received reinforcement directly on every fourth item and only these fifteen items were used in the comparison multiplied by four to make the means roughly comparable.

Summary of Findings

This study was completed in order to investigate learning conditionsinvolving pupil-pairs working together on a verbal learning task and pupils working alone on the same task. Four learning conditions were established; three involving pupil pairs and one with pupils working alone. Subjects were 192 elementary school pupils equally divided between fourth, fifth and sixth grade levels. Subjects were also equally divided across learning conditions. Subjects were paired by reading score and sex. Sets of subjects for each learning condition were equated by reading scores.

The task consisted of sixty German stimulus nouns each with two English responses. The task was to learn the German noun and the one English response which had the same meaning as the German word. The

task was divided into three equal parts and one part was given on each of the days, Monday, Tuesday and Wednesday with a recognition test given on Friday for all sixty items. A delayed test of retention of the task was also given to all subjects.

Time data was collected on all subjects by recording the time each subject or pair of subjects required to complete each day's task. Observations of behavior other than procedural behavior was recorded by the experimenter.

Findings in the study were examined across learning conditions, grade level, and comparative conditions with a previous study. The findings were as follows:

Condition I and II (a pupil pair with one subject [I] in the teacher role and the other subject in the pupil role [II]). Subjects who served as teachers in the pupil pair learned significantly less than subjects in all other conditions. These subjects also had the poorest efficiency scores when time to learn the task was considered. Subjects in Condition II, the pupil role, had significantly superior learning to all other conditions on the immediate test but learning was superior only to Condition I subjects on the delayed test. Efficiency scores for subjects in Condition II were second only to subjects who worked alone.

Condition III (pupil pairs where subjects switched teacher-pupil roles at mid-point in each day's task). Subjects in this learning condition learned significantly more than those in Condition I and showed the least percentage of loss from the immediate test to the delayed test. Poor efficiency scores were recorded. Change of role

appeared to make this condition more motivating to subjects than other conditions.

<u>Condition IV</u> (pupil working alone with self-instruction materials). These subjects were second only to subjects in Condition II on total learning scores. Subjects in this condition learned the first day's task better than any other subjects, but showed next to the lowest learning on the third day's task. This decrement was in part responsible for one double interaction in the analysis of variance being significant. Subjects in this condition had the best efficiency scores requiring less than half the time per item learned than two other conditions.

When learning scores were compared across grade levels no significant differences were found; however, some differences attributable to grade level were evident.

<u>Grade Four</u>. Subjects from this grade level had higher scores in pupil-pairs than when working alone on the task. Efficiency scores for this grade level were poorest of the three levels. Subjects in this grade were the most likely to show behavior other than procedural; these subjects accounted for forty-one per cent of such behavior recorded by the experimenter.

<u>Grade Five</u>. These subjects appeared to be the norm of the three levels. Fifth graders showed good learning in all conditions and had efficiency scores which were better than fourth graders, but not as favorable as sixth graders.

<u>Grade Six</u>. Subjects at this level showed excellent scores when working on the task alone, and were less effective when working in pairs.

Efficiency scores were superior to other grade levels. Observed behavior indicated these subjects had a great desire to conform to peer and experimenter expectations.

Some comparisons made with the Van Wagenen (Chapter II) study, which used the same task, indicated that subjects can learn this type of task as well by working in pupil-pairs or alone as subjects interacting in small groups with an adult.

Implications:

1. Pupils who received verbal reinforcement feedback learned more than those who gave it. Pupil-pairs in a classroom should be organized in such a way that task materials allow them to reinforce one another.

2. Materials for self-instruction in tasks of this type produced efficient learning, but showed marked decrements after the initial learning. This decrement might be reduced by a redistribution of learning trial time. Materials should probably be constructed to give variety to this experience. A teaching machine produced no better learning than that acquired by use of a simple self-instruction card. Efficiency in learning was superior with the self-instruction mode; it required subjects in this mode about half the time to learn an item than was required by pupil-pairs.

3. Grade levels made little difference in total learning on this task, but lower grade level subjects learned more working in pupil-pairs while higher grade level subjects learned more using a self-instruction mode.

4. Elementary pupils learned this task nearly as well by working in pairs or alone as they did by interacting with an adult teacher-experimenter. Teachers might make more effective use of their time by organizing their

classrooms so that the rote memory tasks are taught by one pupil to another or by a self-instruction device.

5. Subjects who reverse teacher-pupil roles when working in pairs showed about equal learning by both members of the pair. Some implications for classroom organization are evident here, but additional data should be sought which would give information on learning when switching takes place at an increased rate; perhaps on every other item.

6. Spacing of learning trials or a different time distribution of these trials might reduce the learning decrement across days of the task. Additional experimentation should be conducted in this area.

References

Edwards, A. L. Experimental design in psychological research. New York, Holt, Winston and Rinehart, 1962.

Winer, B. J. Statistical principles in experimental design. New York, McGraw Hill, 1962.

CHAPTER VI.

EFFECT OF PUPIL-PUPIL REINFORCEMENT IN MISMATCHED PAIRS

In the previous study, the pupils working in pairs were matched with respect to their scores on a reading test. The research workers were inclined to believe that the amount of learning taking place would be influenced by the extent to which the pupils in a pair were or were not of comparable achievement, but limitations had to be placed on the size of the study and only pairs with matched pupils were included. A second study was planned in which pupils mismatched with respect to achievement were included as well as pairs in which the pupils were matched. While one can list a number of ways in which the effects of mismatching or matching influence the learning of the pupil pairs, it is difficult to make a predicition concerning the way in which the over-all effect will operate. In the mismatched pairs, the high achiever functioning in the teacher role might be expected to urge on his low achieving pupil, but there is also the possibility that he might become impatient and irritated with him. When the low achiever is functioning as a teacher he might well pace the task at too low a rate for the efficient learning of his high achieving pupil. Other problems might well be produced by the fact that high achievers do not typically choose to work with low achievers when the choice is up to them and this mismatching might well provide uncongenial work partners with a resulting deterioration in learning. Rather than stating hypotheses, the purpose of this study was to provide an over-all estimate of the total effect of these various factors on learning in situations in which pupils teach pupils and in which the pupils were mismatched with respect to academic achievement.

The previous study also raised a number of questions concerning the capacity of pupils to work alone at different age levels. The data suggests that the 4th grade pupils worked best in pairs, but by the sixth grade the children working alone were as effective as those working together. An interesting question is whether a similar trend continues through the junior high school level with the older pupil becoming better at working alone than in pairs.

Thus, the study reported in this chapter is a continuation of the study reported in Chapter V in that it reproduced some of the conditions of the latter study but at the 7th and 8th grade levels. In addition, the study also explores further the effect of matching and mismatching pupil pairs in terms of academic achievement on the performance of the pupils in a learning situation requiring the cooperation of each member of a pair.

PROCEDURE

Learning Conditions

Condition I. Pupils matched on the basis of grade point average undertook work in pairs with one pupil functioning as the teacher and the other as the learner. This condition refers to the pupil teacher.

Condition II. This is the same as Condition I except that it refers to the pupil learner rather than the pupil teacher.

Condition III. Pupils worked in pairs as in the previous two conditions but they were mismatched with respect to grade point average. This condition refers to the teacher in each pair.

Condition IV. Pupils worked in pairs as under the previous conditions but the pairs were mismatched. This condition refers to the pupil member of the pair.

Condition V. The pupils worked alone using the cards as a self-instructional device.

Subjects

The subjects for this study were derived from English classes in a junior high school in Salt Lake City. Since the policy of the school is to separate pupils for the purposes of English instruction into high level classes and low level classes, pupils had to be drawn from both of these sets of classes. Eight pupils were drawn at a time and taken to a room where the experiment was conducted. The grade-point average for each child was computed for his or her work covering the three previous semesters, but only grades from academic solids were included in the computation. Pupils from the 7th and 8th grades were included in the study, but the two grade levels were kept separate in all experimental treatments. A total of 160 subjects were included in the study, 80 at each grade level and 32 to each of the five learning conditions. The mean grade-point average for subjects in each grade and each condtion are shown in Table 6.01. The mean difference in grade-point average for the matched pupils was 0.16 and the corresponding difference for the mismatched pupils was 1.21.

Task

The task was identical with that used in the previous studies in this series with the pupil pairs working with small cards with thequestion

TABLE 6.01

MEAN GRADE POINT AVERAGES FOR SUBJECTS BY

GRADE, SEX AND CONDITION

Condition	7th Grade Girls	7th Grade Boys	7th Grade Total	8th Grade Girls	8th Grade Boys	8th Grade Total
Condition I Matched Teacher	2.6	2.2	2.4	2.7	2.1	2.4
Condition II Matched Pupil	2.4	2.2	2.3	2.5	1.9	2.2
Condition III Mismatched Teachers	3.1	2.1	2.6	2.6	2.4	2.5
Condition IV Mismatched Pupils	2.6	2.0	2.3	2.9	2.2	2.5
Condition V	2.1	1.6	1.8	1.9	1.7	1.8

on one side and the answer on the other. Learning took place on three successive days,-Monday, Tuesday, and Wednesday - and the children were tested on Friday.

The data fitted a 2 x 3 x 5 repeated measures design. One hundred and sixty seventh and eighth grade students undertook a learning task on three successive days and were assigned to one of five experimental conditions. Ss worked in teacher-pupil pairs in the first four conditions. In condition five, Ss worked alone with the task.

The task consisted of learning 60 German words and their English equivalents. These words were learned in groups of 20 on Monday, Tuesday and Wednesday of each week. Data were gathered from a multiple-choice test given each pair across the five conditions to complete the task.

Administration of Task

The task was administered as in the previous study. One of the experimenters kept a record of the time required for each pair of pupils or for each pupil for the pupils working alone.

RESULTS

The results will be reported first in terms of the scores on the achievement test, and then in terms of the time data.

Retention Data

Data on individual scores from the retention test was initially examined by computing means and standard deviations for each condition, grade and task. These values are presented in Table 6.02. A brief examination of this data shows that the means for Task A are uniformly higher than those for Tasks B and C. In addition, Means for Task B are consistently higher than those for Task C. It is not clear from this table whether there are any uniform differences as a function of grade level.

In order to examine the differences in amount learned as a function of condition, Table 6.03 was constructed. Means for all conditions were higher for Task A than for Task B and higher for Task B than for Task C.

With the exception of identical means (8.5) for Condition I and II on Task B, the pupils from matched pairs and pupils from mismatched pairs did better than their corresponding teacher partners across all tasks (days).

Tables 6.02 and 6.03 showed differences in amounts learned on each task (day) and also differences in amounts learned on each task as a function of condition. In order to examine the significance of these differences the analysis of variance shown in Table 6.04 was computed with three main effects: tasks, conditions, and grade levels.

Main Effects

The main effects which were significant were conditions ($p<.05$) and tasks ($p<.001$). Differences due to grade level were not significant.

TABLE 6.02

MEANS AND STANDARD DEVIATIONS OF RETENTION SCORES FOR GRADE LEVEL, CONDITIONS AND TASKS--CUMULATIVE GRADE POINT AVERAGES FOR CONDITIONS INCLUDED (N=16 CASES FOR EACH GRADE BY CONDITION)

			A		B		C		Total	
Condition	Grade	CGPA	$\bar{X}$	Σ	$\bar{X}$	Σ	$\bar{X}$	Σ	$\bar{X}$	Σ
I	7	2.4	8.9	2.89	8.5	3.63	8.0	3.33	25.4	3.12
	8	2.4	10.3	2.86	8.5	3.84	6.8	2.27	25.6	3.22
II	7	2.3	11.7	3.71	8.3	3.29	7.8	2.69	27.8	3.76
	8	2.2	11.7	3.03	8.7	3.35	8.3	3.09	28.7	3.24
III	7	2.6	9.6	3.76	8.1	2.74	7.0	4.24	24.7	3.08
	8	2.5	11.3	4.41	9.5	4.87	7.8	2.93	28.6	3.10
IV	7	2.3	12.9	3.67	10.0	3.61	8.4	2.90	31.3	3.41
	8	2.5	12.1	3.65	9.8	3.06	9.5	3.41	31.4	3.49
V	7	1.8	12.0	3.54	9.0	2.32	6.9	3.23	27.9	3.71
	8	1.9	11.0	3.57	8.8	3.05	6.9	2.72	26.7	3.52
Mean and Standard Deviations for Tasks			11.1	3.26	8.9	2.50	7.7	2.98	27.7	

TABLE 6.03

MEANS AND STANDARD DEVIATIONS OF RETENTION DATA FOR TASK BY CONDITION WITHOUT REGARD TO GRADE LEVEL

	A		B		C		Total	
Condition	$\overline{X}$	Σ	$\overline{X}$	Σ	$\overline{X}$	Σ	$\overline{X}$	Σ
I	9.6	2.72	8.5	3.59	7.4	2.76	25.5	3.15
II	11.7	3.19	8.5	2.97	8.0	2.78	28.2	3.33
III	10.8	2.83	9.1	2.85	7.4	2.98	27.3	3.17
IV	12.5	3.49	9.9	3.13	8.9	3.15	31.3	3.56
V	11.6	3.53	8.9	2.66	6.9	2.93	27.4	3.60
Means and Standard Deviations by Task	11.2	3.26	8.9	2.50	7.7	2.50		

TABLE 6.04

ANALYSIS OF VARIANCE OF RETENTION TEST SCORES FOR

TASKS, GRADES AND CONDITIONS

	Ss	df	Ms	F	p
Between Ss	2741.50	159			
Grades	11.72	1	11.72	.71	N.S.
Conditions	197.95	4	49.49	3.00	.05
Grades x Conditions	56.64	4	14.16	.86	N.S.
Residual Between Ss	2475.19	150	16.50		
Within Ss	1638.00	320			
Tasks	1013.68	2	506.84	312.86	.001
Tasks x Grades	.71	2	.36	.22	N.S.
Tasks x Conditions	80.96	8	10.12	6.25	.001
Tasks x Grades x Conditions	55.53	8	6.94	4.28	.001
Within Residual	487.12	300	1.62		
Total	4379.50	479			

As Table 6.05 indicates, all Ss in Condition II (Ss acting as pupils) learned significantly more than did their corresponding teacher partners (Condition I). It is also apparent that the pupils from mismatched pairs in Condition IV learned a significantly greater amount than did Ss in any other condition.

Interactions

The analysis of variance of the retention data yielded a double and a triple interaction which will be examined in that order.

The double interaction was between Tasks and Conditions and was significant at the .001 level. Figure I represents this interaction. It can be seen that the interaction was largely due to the sharp decrement across days in Condition III, teachers from mismatched pairs, and Condition V, pupils working alone with the task.

The triple interaction in Table 6.04, Task x Grade x Condition was significant at the .001 level. Because it is generally agreed that the source of variation in a triple interaction is difficult, if not impossible to represent, no analysis of this interaction was attempted.

Time Data

The time data was obtained by recording the length of time taken by each pair or individual within a condition to complete each day's task. (Example: mean number of minutes for a pair on a particular day's task was 10 3/4 minutes).

Table 6.06 presents the means and standard deviations for amount of time taken per task, condition and grade level. It should be noted

TABLE 6.05

DUNCAN'S MULTIPLE RANGE TEST FOR SIGNIFICANCE OF DIFFERENCE IN AMOUNT LEARNED BY EACH CONDITION

Condition	I	III	V	II	IV
I	--	N.S.	N.S.	.05*	.01**
III		--	N.S.	N.S.	.01
V			--	N.S.	.01
II				--	.01
IV					--

* The .05 refers to the level of significance of the difference between Condition II and I ($p<.05$, Condition I < Condition II).

** The .01 refers to the level of significance of the difference between Conditions I, III, V and II ($p<.01$, Condition I, III, V and II< Condition IV).

N.S.=Not significant

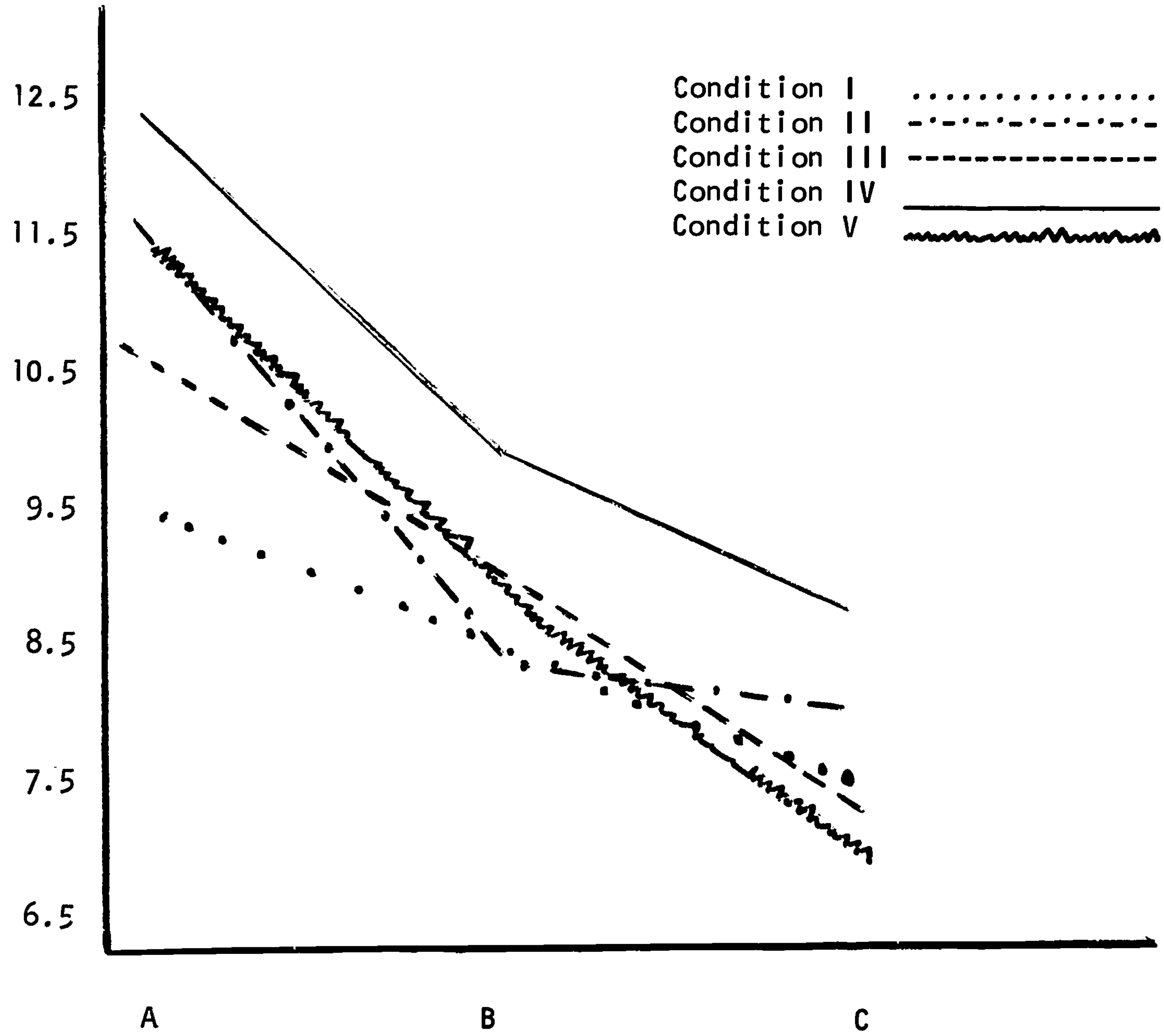


Fig. 6.01 Mean retention scores demonstrating double interaction of tasks x conditions for retention data.

TABLE 6.06

MEANS AND STANDARD DEVIATIONS OF TIME DATA FOR

GRADE LEVEL, CONDITION AND TASK

Condition	Grade	A		B		C		Total	
		$\bar{X}$	Σ	$\bar{X}$	Σ	$\bar{X}$	Σ	$\bar{X}$	Σ
I	7	10.70	1.92	10.73	1.88	9.67	2.57	31.10	2.16
	8	11.70	3.08	11.44	2.54	10.16	2.10	33.30	2.64
II	7	10.70	1.92	10.73	1.88	9.67	2.57	31.10	2.16
	8	11.70	3.08	11.44	2.54	10.16	2.10	33.30	2.64
III	7	12.09	1.66	10.82	1.74	9.61	1.92	32.52	2.01
	8	13.46	1.89	12.53	1.64	11.89	2.16	37.88	1.98
IV	7	12.09	1.66	10.82	1.74	9.61	1.92	32.52	2.01
	8	13.46	1.89	12.53	1.64	11.89	2.16	37.88	1.98
V	7	6.77	1.68	6.75	2.26	5.77	2.55	19.29	2.20
	8	6.46	2.68	5.36	2.16	4.76	1.72	16.58	2.28
Means and Standard Deviations for Tasks		10.91	3.17	10.31	2.99	9.31	3.07		

on this and the following representations of the time data that the figures presented indicate the amount of time in minutes taken to complete one day's task under a particular condition. (Example: seventh grade, Condition I, Task A, time taken was 10.70 minutes). It should further be noted that the time data for Conditions I and II are identical as are those for Conditions III and IV. This is inevitable since Condition I and II represent a matched pupil-teacher pair and both teacher and pupil have to take the same amount of time to complete the task, and Conditions III and IV represent a mismatched pupil-teacher pair and, again, both members of each pair take the same amount of time.

A brief examination of this data shows that subjects across all five conditions took uniformly more time on Task A (first day) than they did on Task B or C (second and third day). They also took more time on Task B than on Task C.

An examination of time differences by condition shows that Conditions III and IV (mismatched pupil-teacher pairs) took a uniformly greater amount of time across all tasks than did Ss in Condition I and II. Condition V (pupils working alone with the task) took less time across all tasks than did any of the other Ss.

The table also shows that in the paired conditions the eighth grade Ss consistently took a longer period of time across all tasks than did the seventh grade Ss. The exception to this situation was Condition V in which the eighth grade Ss took a shorter time each day than the seventh grade Ss.

In order to examine the significance of the time differences in tasks, conditions, and grades the analysis of variance shown in Table 6.07 was computed with the above mentioned three main effects.

Main Effects

Time differences across the three tasks were found to be significant at the .001 level. A brief examination of Table 6.06 shows that the variance in the means of the five conditions can be accounted for almost entirely by the fact that Ss in Condition V took approximately half as much time to complete all three tasks as did Ss in the paired conditions.

The differences in time taken within conditions were significant at the .001 level. A Duncan Multiple Range Test was computed to determine the significance of these differences. This data is presented in Table 6.08. It can be seen that Ss in both paired conditions took a significantly longer time with the task than did Ss in Condition V who worked alone with the task. The data further indicated that Ss in Conditions III and IV (mismatched teacher-pupil pairs) took significantly longer on each day's task than did Ss in Conditions I and II (matched pupil pairs.

Interactions

One double interaction between grade and learning conditions was found to be significant at the .05 level. This interaction is represented in Figure 6.02, but it does not appear to be particularly meaningful.

Efficiency of Learning

One of the purposes of this study was to replicate certain conditions which were examined in the study reported in Chapter V. In the

TABLE 6.07

ANALYSIS OF VARIANCE OF TIME DATA FOR

TASKS, GRADES AND CONDITIONS

	Ss	df	Ms	F	p
Between Ss	3955.53	159			
Grades	81.09	1	81.09	7.90	.01
Conditions	2217.84	4	554.06	53.95	.001
Grades x Conditions	116.72	4	29.18	2.84	.05
Residual Between Ss	1539.88	150	10.27		
Within Ss	773.50	320			
Tasks	207.41	2	103.10	57.20	.001
Tasks x Grades	1.12	2	.56	.31	N.S.
Tasks x Conditions	18.08	8	2.26	1.25	N.S.
Tasks x Grades x Conditions	12.32	8	1.54	.85	N.S.
Within Residual	534.57	300	1.81		
Grand Total	4729.03	479			

TABLE 6.08

DUNCAN'S MULTIPLE RANGE TEST FOR SIGNIFICANCE OF DIFFERENCES IN TIME TAKEN BY CONDITIONS

Condition	V	I	II	III	IV
V	--	.01*	.01	.01	.01
I		--	--	.01	.01
II			--	.01	.01
III				--	--
IV					--

*The .01 refers to the level of significance of the difference between Conditions I, II, III, IV and V.

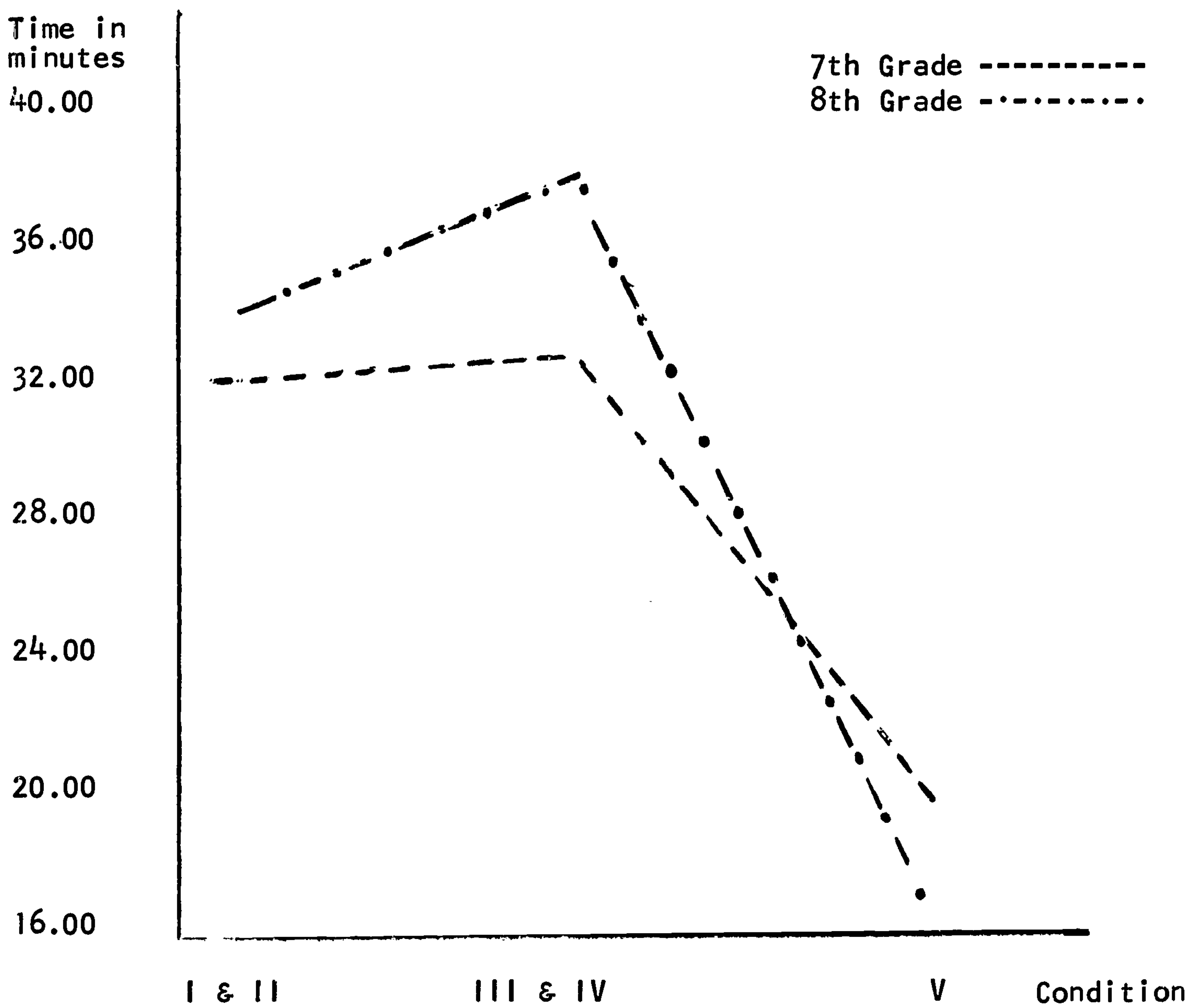


Fig. 6.02 Mean time across days representing double interaction of conditions and grades for time data.

study it was stated that a basic reason for recording the time data was to attempt to develop a measure of learning efficiency. This same rationale was also used in recording time data in the present study. A learning efficiency score was developed for each person in each condition and each grade level from the formula $ES = \frac{TT}{CR}$ (ES = Efficiency Score, TT = Total Time, CR = Correct Responses on Retention Test). The fewer minutes required per correct response learned, the more efficient the learning was considered to be.

The formula was applied to all cases in the five conditions and the summary data are provided in Table 6.09. Condition V, self-instruction, yielded the most efficient learning with a score of 0.65. The other four conditions required approximately double that time to learn a correct response.

Summary data for the efficiency scores by grade level are presented in Table 6.10. Although the difference in efficiency score by grade level is slight, eighth graders were slightly more efficient and seventh graders slightly less.

Comparison of Learning Conditions in this Study with Comparable Conditions in the Previous Study

One of the main purposes of this study was to reproduce three of the conditions of the previous study (1964) but with older children and to see whether any marked trends over age could be identified. The three conditions which were duplicated for comparison were Condition I (teachers from the matched teacher-pupil pairs), Condition II (pupils from the matched teacher-pupil pairs) and Condition V (pupils working alone). The

TABLE 6.09

LEARNING EFFICIENCY SCORES FOR SUBJECTS IN

CONDITIONS I, II, III, IV AND V

Condition		Total Time	Correct Response	Time per Item
I	(Matched teachers)	1030.50	817	1.26
II	(Matched pupils)	1030.50	907	1.13
III	(Mismatched teachers)	1126.50	877	1.28
IV	(Mismatched pupils)	1126.50	1006	1.11
V	(Isolation)	573.90	882	.65

TABLE 6.10

LEARNING EFFICIENCY SCORES FOR SUBJECTS IN

GRADES SEVEN AND EIGHT

Grade	Total Time	Correct Response	Time per Item
7	2345.30	2207	.941
8	2542.68	2282	.897

data from these conditions which was compared were the number of correct responses, the time data and efficiency scores. The previous study used fourth, fifth and sixth graders in a public elementary school. The present study used seventh and eighth graders in a public junior high school

Figure 6.03 presents the mean number of correct responses for the three conditions by grade. Although it is difficult to describe any definite trend, it can be seen that subjects who acted as pupils (Condition II) appeared to learn somewhat more across all grades than did subjects in any other condition. The graph also clearly indicated that subjects in Condition I (matched teachers) continued to learn less across all grades than did their pupil counterparts. The result of these two trends is for the gap between teachers and pupils to close with increasing age.

Figure 6.04 presented the mean number of correct responses for the three conditions and the five grades studied. As is indicated, subjects in the fourth, fifth and sixth grades seemed to learn a slightly greater number of correct responses than did subjects in the seventh and eighth grades. The effect is undoubtedly motivational.

The time taken to complete the task by grade level is shown in Figure 6.05. The time shows a decline across grades, but the decline does not correspond to an increase in efficiency. The data suggest that the pupils in the higher grades tend to dispose of the task more rapidly but that they could spend more time on it with a corresponding increment in learning.

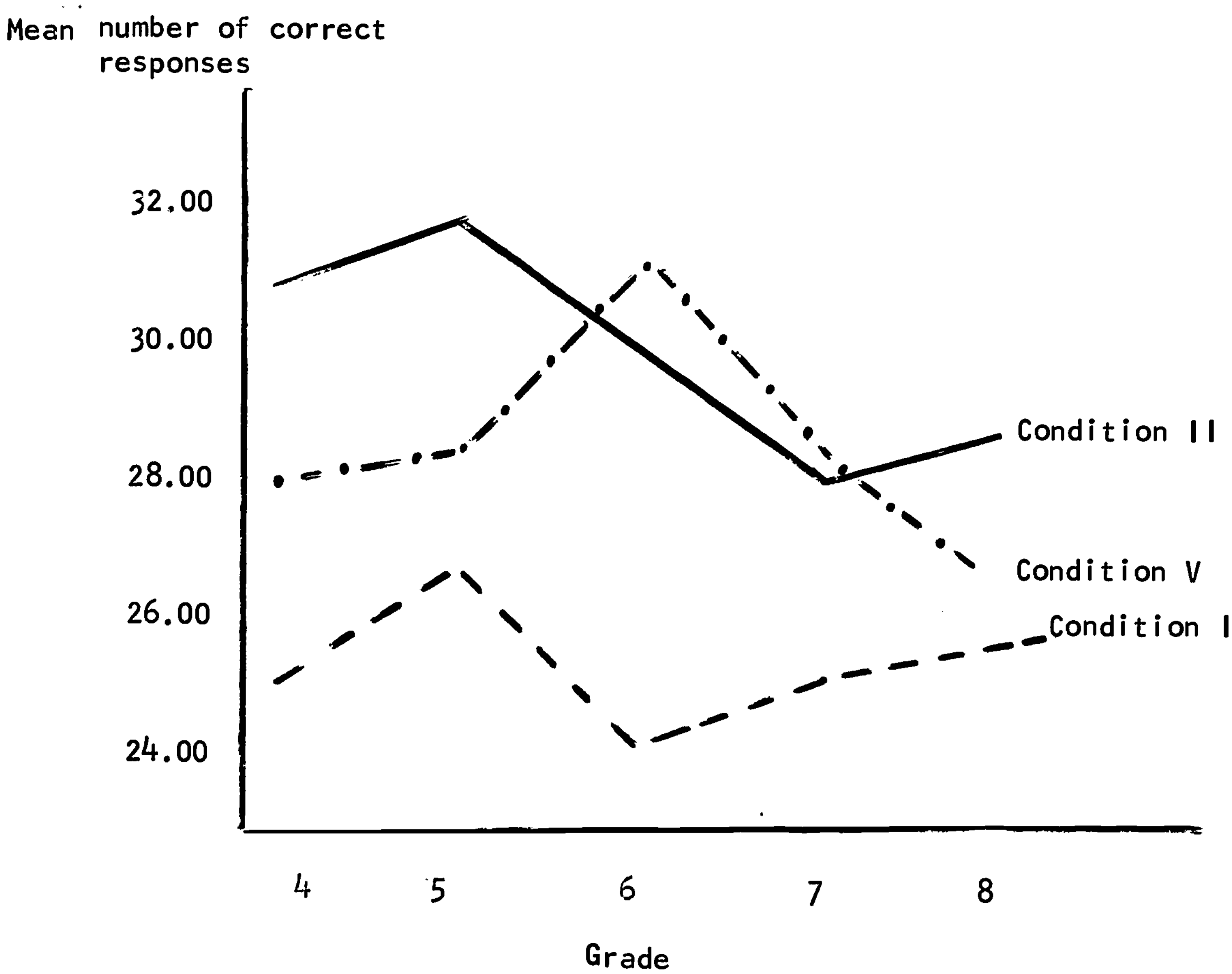


Fig. 6.03 Comparison by grade and condition of mean number of correct responses in this study and the previous study.

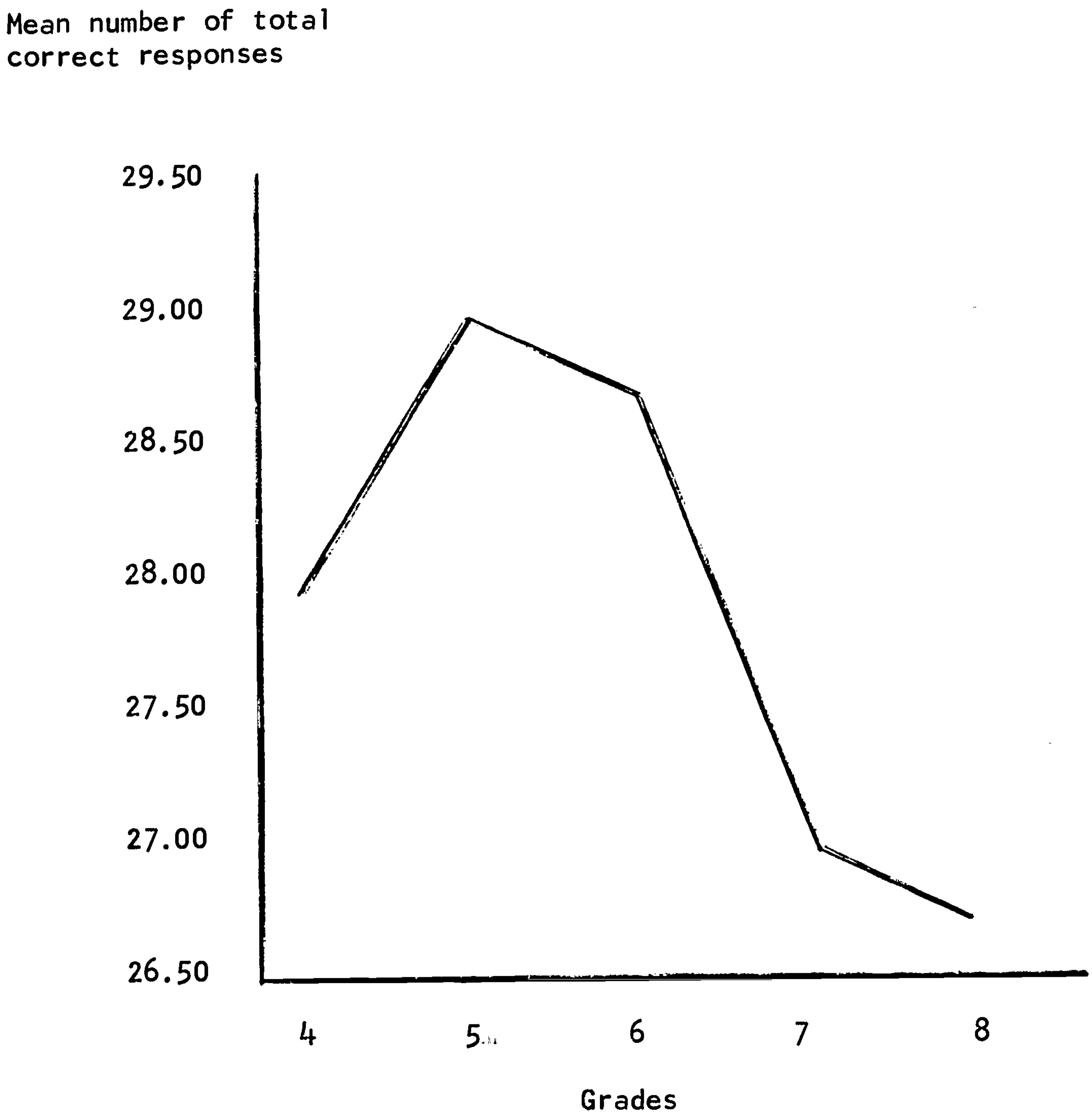


Fig. 6.04 Mean number of correct responses across Conditions I, II and V for the five grades which were compared.

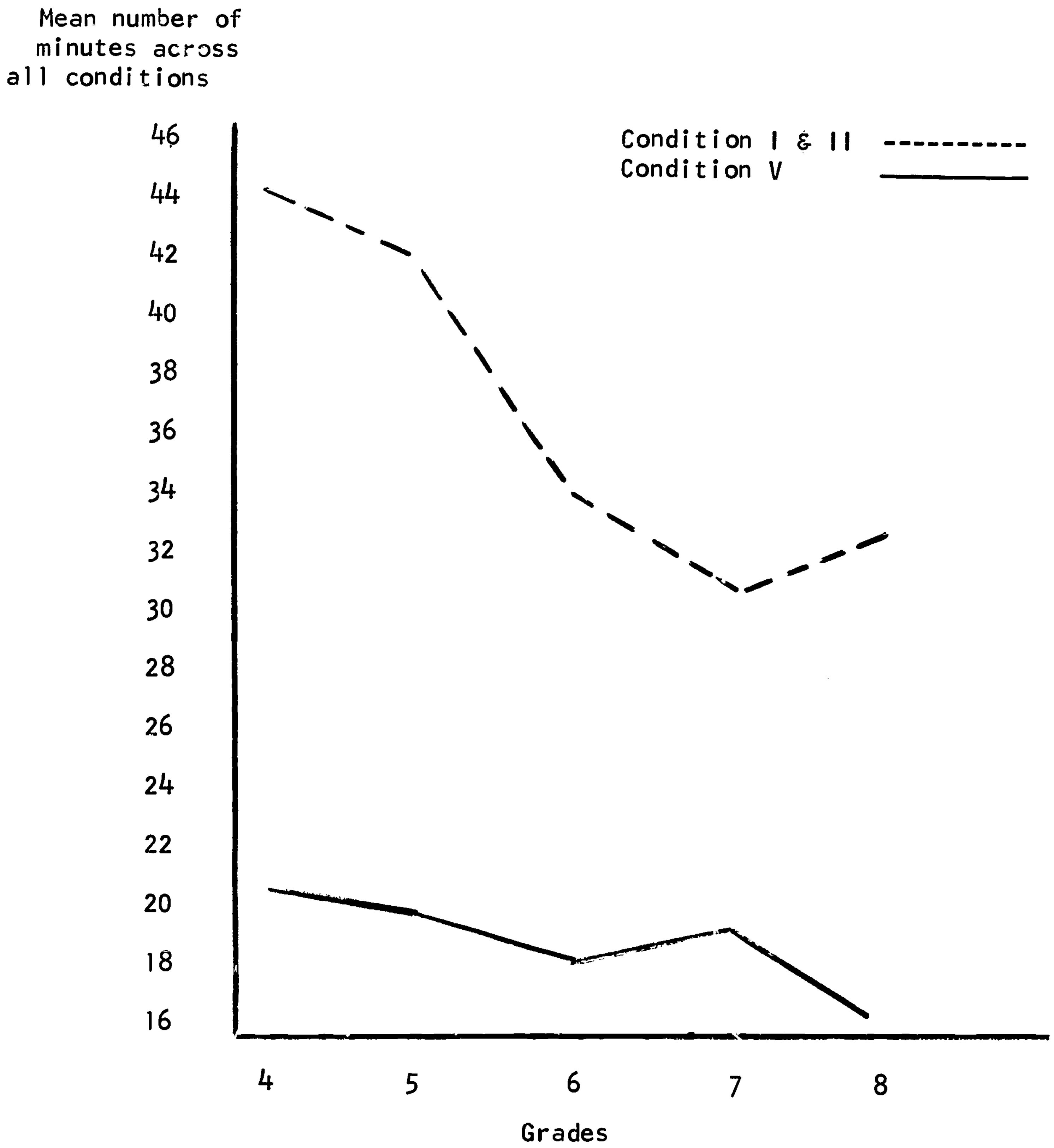


Fig. 6.05 Comparison of mean number of minutes across all tasks in this study and the previous study.

Efficiency scores across the five grades are presented in Table 6.11. It can be seen that Ss in the seventh and eighth grades generally took a shorter time, but the time per item learned was suprisingly constant. The higher grades learned fewer items to the point of correct performance on the test, a fact which probably reflected poor motivation in the experimental situation at the junior high school level. The elementary school pupil cooperates in experiments to a far greater extent than the more blasé seventh or eighth grader.

DISCUSSION

The study raises a number of questions of importance to the management of classroom learning.

First, the differences in the performance of the matched and mismatched pupils favors the mismatched group. This needs to be investigated further before any attempt is made to apply the result to educational practice. The expectation had been that the pairs containing matched pupils would provide the more effective learning situation, but the reverse was the case. A number of observations made by the experimenters give leads concerning the reasons for the differences found. The experimenters observed that in the mismatched groups there seemed to be more interaction between the pupil functioning as a pupil and the pupil functioning as the teacher. The high grade-point teacher tended to be more reprimanding when the pupil made a wrong answer than the teacher in the matched pairs. The data also indicate that the high level of performance of the mismatched pairs was a result of the high performance of the high grade-point pupils in the pupil role.

TABLE 6.11

COMPARISON OF LEARNING EFFICIENCY SCORES ACROSS THREE COMPARABLE CONDITIONS FOR SUBJECTS IN GRADES 4, 5, 6, 7 AND 8

Grade	Total Time	Correct Response	Efficiency Scores Time per Item
4	1785.8	1342	1.33
5	1674.6	1398	1.10
6	1379.1	1375	1.00
7.	1304.5	1268	1.02
8	1330.4	1302	1.02

Second, the data in this study also support the data previously presented indicating that while the individual working alone is the most efficient as a learner, his decrement over days is much greater than the decrement found under other learning conditions. One is left wondering what would have happened if the task had been spread over several weeks with learning sessions on each day. One suspects that the social conditions would then have shown themselves to be considerably more efficient than the condition which did not involve social interaction. The latter has been the experience of those who have studied programmed learning and teaching machines. Such studies have suggested that the first reaction to a teaching machine is highly effective learning, but soon the novelty falls off and a very marked decrement in learning occurs. Our data suggest that the endless novelty found in social interactions may have the effect of keeping the learner aroused and hence working at least at a moderately efficient level.

Third, the fact that there is no particularly marked increment in performance across the five grades studied is, at first sight, surprising. Most tasks that have been studied do show a very marked increment. The lack of any increment in performance across grades with the present task may be a result of the fact that level of performance on the task is probably highly dependent upon motivational factors. The task is certainly within the ability of all of the pupils who were exposed to it, if they would set themselves to learn it. Differences in performance between Conditions II, IV and V all of which involve a response-reinforcement paradigm must be essentially motivational in character.

Finally, the point must be made that in this study, as in the other studies involved in this series, those learning conditions which require the learner to respond and which then follow the response with some form of reinforcement are generally superior to those that depart from this learning formula.

CHAPTER VII

THE ABILITY OF PUPILS TO USE POSITIVE AND NEGATIVE INFORMATION DERIVED FROM OBSERVING THE BEHAVIOR OF OTHER PUPILS

Study I

In the study reported here an attempt is made to determine the extent to which "observer pupils" obtain information specifically from the behavior of those "pupils" who interact and to discover some of the conditions that influence the learning of the "observer pupils." The paradigm followed in the study was that of the teacher presenting the pupil with a problem and a two-choice answer. The interacting pupil chooses one of the answers and is told if he is right. Whichever choice the interacting pupil makes, he provides an equal amount of information to the observer pupil. That is to say, the observer pupil can always determine the right answer from the behavior of the interacting pupil. However, there is considerable interest in determining whether the observer pupil learns as much when the interacting pupil is correct as when he is incorrect.

Of particular importance in the present study is the extent to which a pupil learns from the errors as contrasted with the correct responses made by other pupils. Evidence from the studies of Hovland and Weiss (1953) and Bruner, Goodnow, and Austin (1956) indicates that, when an error or a correct response provides equal information in a concept learning task, subjects learn more from correct responses than from errors. Whether a similar effect occurs when one pupil observes the behavior of another is a matter to be determined.

In the experiment reported in Chapter II, observer Ss were under the influence of varying efficiencies for direct subject oral responding. That is, they heard direct Ss respond correctly in some instances to all four presentations of some items, while for other items, 0, 1, 2, or 3

trials were observed by them to be correct. The purpose of that part of the investigation reported here was to determine the extent to which the efficiency of this oral responding in the learning trials influenced the retention of correct responses by observer Ss who had the experience of observing the varying efficiencies during the trials. To be borne in mind is the fact that knowledge of results was afforded after each item had been responded to, so that the same information was available to the observer Ss whether a direct S responded correctly or not. A comparison of the retention of particular items by observer Ss (as measured Friday) with the recorded learning-trials performance of direct Ss on those same items was undertaken. An example may suffice to clarify the procedure: assume a hypothetical direct subject X. Let us say that subject X was confronted on a Monday (or Tuesday, or Wednesday) with the item "Wahrheit" to which he gave audible responses. This item, as was the case with all other items, was presented four times at irregular intervals. Assuming that subject X gave two correct oral responses to the item and two incorrect ones, one might suspect that the four observer Ss observing this responding efficiency would be uninfluenced by the number of correct performances since the correct choice was always indicated by the experimenter. On the other hand, if the observer Ss attended more closely to the responding of their peers than to the feedback of the experimenter, learning might be expected to be less effective for them if the direct S made four incorrect choices.

Results

The learning of the observer subjects cannot be traced as it progressed on particular items through the first three days of the week since

the only measure of learning obtained from them was administered on Friday, However, the course of learning for the direct subjects can be traced item by item as it occurred in the simulated classroom situation, and retention can also be measured on Friday. Since each item was presented to the same direct Ss on four different occassions during an acquisition session, the correctness of the response on each trial can be determined. One would expect an increasing probability of correct responses over the four trials.

On every presentation of every item, the card showing the two-choice problem was clearly visible to all Ss, including those involved only as observers. The correct answer could always be determined by observing the behavior of the direct S and E. That is, the observer S could see the direct subject respond either correctly or incorrectly and E present knowledge of results. On each item he would have opportunity to observe the direct subject responding correctly between 0 and 4 times. The problem is to determine whether the learning of the observer subjects is related to the number of times they observe a correct response, even though each trial provides all the information necessary for learning. Now it is possible to find items that were responded to correctly by the direct Ss 0 times out of 4 trials, 1 trial correctly, 2 trials correctly, etc. From the data it is then possible to determine the probability that an observer S would respond correctly to the item on Friday under each one of

these five learning conditions. These probabilities are presented in Figure 7.01 (top).

The data show an increasing probability of observer Ss answering correctly depending on the number of correct responses to which the indirectly involved subject had been exposed.

Figure 7.01 (bottom) shows the same data as for the upper graph but broken down according to whether the items were learned on Monday, Tuesday, or Wednesday. They show a positive slope for each day of learning and also the fact that there is declining retention across these three days. The latter effect may be interpreted either as a result of proactive inhibition or as a result of attention becoming increasingly erratic.

Other explanations of the effect noted need to be investigated. First, there is the possibility that although the items were constructed to be equal in difficulty, there might still be substantial differences. If the latter were so, then the items which the direct subjects missed most frequently in the recitation situation would also be those which the observer subjects would be most likely to miss on Friday's test. In order to evaluate this explanation, an item analysis was made for an independent group of 200 who were exposed to a similar learning procedure and who took the test also on Friday. For this group a comparison was made of the performance on the test of those items that had been answered zero times by the direct Ss in the present study with those items that had been answered 4 times correctly. The probability of the first group of items (0 presentations correct) being answered correctly on the Friday test was 0.554, while the probability of the latter group (4 presentations correct) being answered correctly was 0.615. The difference between these two

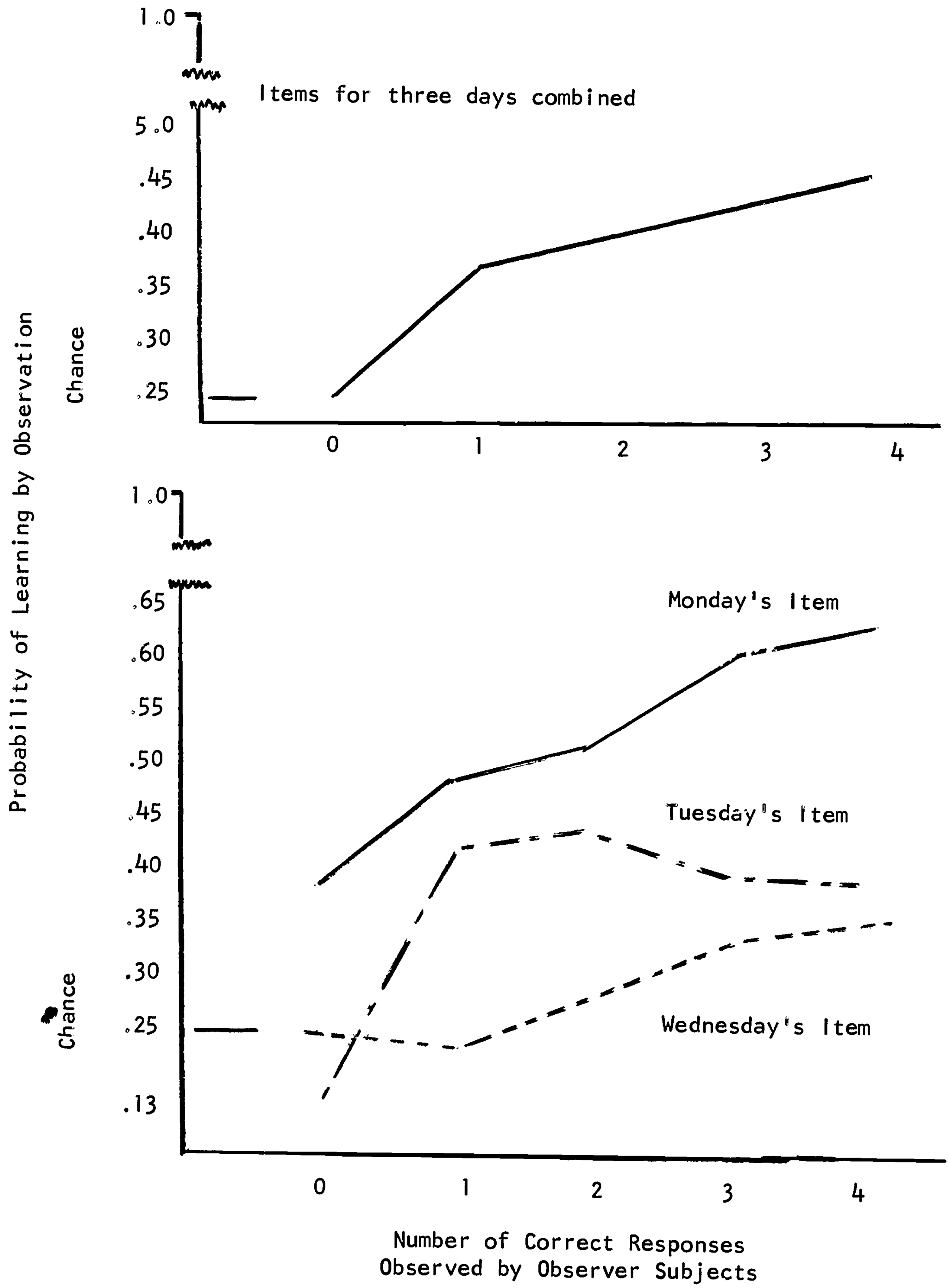


Fig. 7.01. Probability of learning by observation at five levels of accuracy for the subject being observed.

probabilities is only 0.061, a difference very small compared with the difference found in the case of the observer Ss who had seen the one group of items responded to by direct Ss 0 times correctly and the other group of items 4 times correctly. The weight of the evidence supports the position that observer Ss learn more when they see another subject respond correctly than when they observe him respond incorrectly.

The small differences in difficulty of the items on the Friday test might be accounted for in terms of slight differences in initial association value, although every effort had been made to eliminate this factor. That this factor is small is reflected by the fact that on the first trial the direct subjects would have chosen the correct answer with a probability of 0.500. In actual fact, the data showed a slight positive association value, with the correct answer being guessed with a probability of 0.512. Nevertheless, there appear to be some small consistent differences. When such a probability was computed for each item for the direct Ss and a similar probability was computed for the so-called control group, the two sets of probabilities correlated C.27 (p=.0.05).

Discussion

Evidence has been presented to indicate support for the position that pupils who observe other pupils interacting with a person in the role of teacher learn more from the correct responses of the interacting pupil than from his errors. This occurs despite the fact that, in the situation studied errors and correct responses carry equal information. The situation parallels closely that found in concept-formation studies in which Ss performing tasks related to the acquisition of concepts obtain more

information from their correct responses than from their incorrect ones.

The implication of the findings for teaching is that the teacher should not let the correct response be known only by implication. Correct answers should be clearly stated and pupils should not be expected to infer a correct answer from the errors made by other pupils.

Study 2

The previous study, which was based on data derived from another experiment suffered from the defect that it did not provide a means of controlling for differences in association value of the different items. The phenomena involved appeared to be of sufficient interest to merit the undertaking of a separate experiment which would have as its central purpose the exploration of the effect of hearing the right answer a given number of times and the wrong answer a given number of times in a learning situation in which both right and wrong answers provide equal amounts of information. In this second experiment an advantage was to be gained by the fact that right and wrong answers were to be given by stooges who, over the series of learning trials, could be cued to give the wrong answer either zero times, or once, or twice, or three times or four times. The experiment could be so arranged that the items for which the correct answer was given zero times would, on the average, be equal in association value to the items for which the correct answer was given once. The same applied to the items to which a correct answer was given twice, three times and four times. Observer Ss were to learn by observing the performance of the stooges as these responded correctly to items zero, once, twice, three times, or four where the proportion of items was equal under the five categories.

Method

Materials. The tasks which were used in the previous experiment were used in this study, but with some rearrangement of the materials. Task C was given on the first day, and was followed on subsequent days by Task A and B. This change in the order of the tasks served another purpose and need not be considered here. However, an important rearrangement of the items within each task was made. The purpose of this rearrangement was to classify the items into five groups so that the average association values for each group was the same. The basis for this classification was the extent to which subjects in the experiment reported in Chapter II had been successful with individual items in the task. According to the criterion performance of these subjects, the items were ordered in descending difficulty and then assigned to the categories of 0, 1, 2, 3 and 4 by which stooges were to respond during the learning trials. Items in any group were to be answered by the subjects by a given number of right and wrong answers over the four trials. The order of correct and incorrect responding was programmed so as to provide a like-random order of responding to items under the five categories. To illustrate this, the order of correct responding for item number 1 could have been right-wrong-right-wrong and for item number 2, it could have been wrong-wrong-right-wrong. All of the possible combinations of correct responding were employed in the task. Each group of items was assigned to such a response category before the experimental data were collected.

Subjects. All the subjects (N=180) came from grades 4, 5 and 6 in two elementary schools of Tempe, Arizona, a University community of 40,000 inhabitants. There were six classrooms representing three grade levels at

each of the two schools. The four subjects from a school classroom who were to be those responding to the items and who were to be the stooges in the experiment were separated from the class and instructed that they were going to help in the conduct of an experiment and that they were to perform exactly as instructed. These pupils were selected by the teacher as those most likely to cooperate. They were shown the general nature of the task and were told that when the experimenter's thumb appeared in front of the card that they were always to choose the first answer and when they could not see his thumb in front of the card that they were to choose the second answer. Each one of the four stooges was then given a practice trial to make sure that he understood the instructions. They were also told that they must not tell any of the other children in the school that they had been given special instructions. Classroom teachers made the selection of stooges (two girls and two boys) on the basis of the teacher's expectation of whom among her pupils could be relied upon to keep a secret. Care was taken to ensure that none of the subjects had any familiarity with German. Each group consisted of 4 stooges and approximately 25 pupils who learned by observation.

Task administration conditions. The administration conditions involved a classroom situation in which the children were seated in their normal desk positions. The names of the four stooges were called and these took positions in a straight row in front of the class at a 45° angle to the class on the right and the experimenter on the left. The experimenter also stood in front of the class with the large cards presenting the German words. The pupils learned the task at the same hour each day on Monday, Tuesday and Wednesday, and were tested in their regular classrooms

on Friday. Both the stooges and the other pupils were told that they would be tested sometime for their knowledge of the German words presented. The feedback conditions were identical with those provided in the first study reported in this series. When the correct answer was given the experimenter said "right" or "that's right." When the pupil gave the wrong answer the experimenter said nothing. The task was administered only to the stooges. The remainder of the pupils learned the task by observation

Criterion task. The same test which was given in previous experiments was given on Friday to all those participating in the learning sessions.

Treatment of the data. When a check was made on the difficulty level of the items that had been assigned to the five different learning conditions (observing 0 correct responses, 2 correct responses, etc), it was found that the groups of items did not have the equality which they were supposed to have. In order to provide for the equality of the groups of items, an essential condition of the experiment, an item was discarded from each group of five items, leaving four items in each learning condition. The discarding of the items was undertaken after the data had been collected, but the procedure introduced no bias into the results since it was based on data collected in an earlier study.

The test given on Friday was then scored for three groups of 15 items, and not for three groups of 20 items as was the case in the other experiments reported in this series.

Results

The means and standard deviations relevant to the analysis are presented in Table 7.01. There is certainly no trend comparable to the

TABLE 7.01

MEANS AND STANDARD DEVIATION FOR SCORES ON

ACHIEVEMENT TEST ADMINISTERED FRIDAY

		0 right	1 right	2 right	3 right	4 right	Means by Days
Monday	M	1.46	1.12	1.01	1.05	1.32	5.96
	SD	.83	.88	.85	.85	.86	
Tuesday	M	1.46	1.38	1.57	1.17	1.15	6.72
	SD	.90	.87	.95	.92	.85	
Wednesday	M	1.09	0.99	1.30	1.22	0.95	5.55
	SD	.94	.86	.86	.87	.90	
Means by number of right responses observed		3.99	3.49	3.88	3.44	3.42	

trend evident in the previous study. Indeed, if there is any trend at all it is in the reverse direction with the greater amount of learning occurring under the condition where the right answer was observed 0 times during the series.

An analysis of variance was undertaken in order to test whether the five different learning conditions had a differential effect on learning. The analysis of variance is shown in Table 7.02. Signfiicant differences were found between days, between learning conditions, and for the interaction of days and learning conditions. An examination of the table of means to determine, if possible, the nature of the significant effects, is not particularly helpful. The lack of orderly arrangement of the means in terms of magnitude suggests that the main effect might be due to differences between item groups which did not appear in the data on the basis of which the items were assigned to learning conditions. For some unknown reason, this particular group of subjects may have found certain groups of items more easily learned than others and the main effect of learning conditions may have been produced by an uncontrolled condition of this kind.

The interaction is presented in graphical form in Figure 7.02. The graphical representation suggests that on the first day of learning, Monday, that the items which were answered correctly every time or zero times were the most readily learned, while on the second and third day the items answered correctly once, twice, or three times were the most readily learned. The interaction may represent a genuine effect and reflect a tendency for subjects to learn how to learn this particular task.

TABLE 7.02

ANALYSIS OF VARIANCE FOR TEST SCORES OF PUPILS

LEARNING BY OBSERVATION

Source	df	Ss	Ms	F	p
Between Subjects	148	197.08			
Within Subjects	2086	1592.4			
Between Days	2	20.93	10.47	14.30	<.01
Between Learning Conditions (0-4 right)	4	14.46	7.23	9.90	<.01
Days X 0-4	8	39.55	4.94	6.75	<.01
Residual	2072	1517.3	.73		

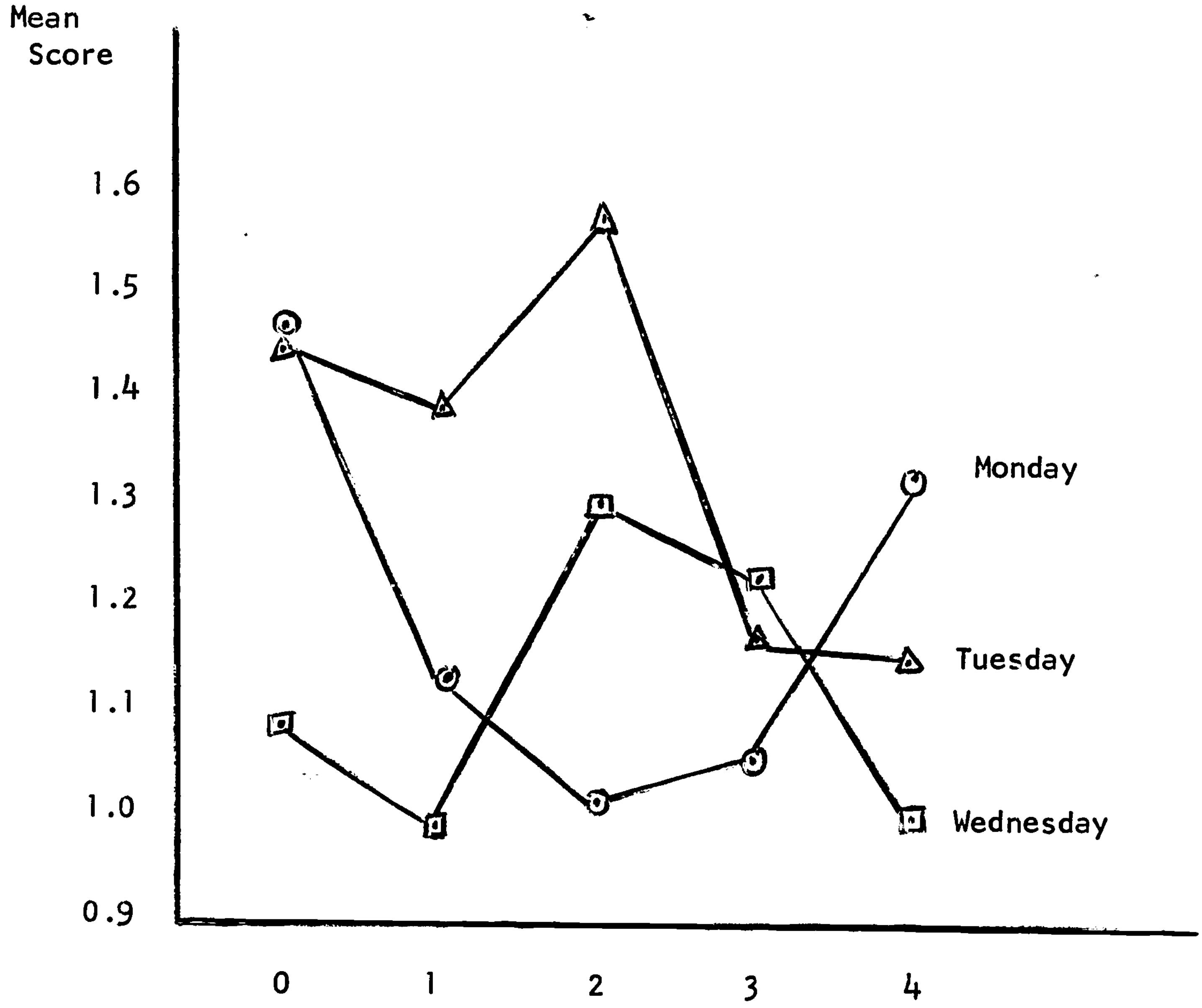


Fig. 7.02 Graphical representation of interaction from Table 7.02.

The Learning of the Stooges

The data derived from the stooges is of some interest in that, while they responded to the situation, the response made was as often wrong as right. Table 7.03 shows the means and standard deviations for the group of 23 stooges tested with the other pupils on the Friday of the week during which the task had been undertaken.

The means show a systematic variation from the items that were responded to zero times correctly to the items that were responded 4 times correctly. The items from which the greatest amount was learned were those to which the stooge sometimes responded correctly and sometimes incorrectly. The mere fact that a correct response was made and that the experimenter said "that's right," does not in itself make for a favorable learning situation. While the stooges had the same amount of information necessary for learning as did the pupils who learned by observation, they learned considerably less. When the means given in Table 7.03 are compared with the means given for the pupils learning by observation, they are found to be in all five instances at a lower level. The sign test involved in this comparison provides a significance of the differences at the .03 level.

The data on the stooges give support to those who doubt the value of the response-reinforcement paradigm for planning or understanding the verbal learning of children. The reinforcement position would lead one to expect that the group of stooges that made the correct response four times and who were reinforced for so doing would learn the task the best. In actual fact they did not and the clear-cut trend is for the best learning to occur when some overt incorrect responses were deliberately made.

TABLE 7.03

MEANS AND STANDARD DEVIATIONS FOR SCORES ON ACHIEVEMENT TEST ADMINISTERED FRIDAY FOR THE STOOGES (N=23)

	0 right	1 right	2 right	3 right	4 right
M	2.65	3.30	3.52	3.00	2.91
SD	1.07	1.63	1.65	1.62	1.50

An Overview of Both Experiments

The studies have not been able to produce any consistent evidence concerning the capability of pupils to benefit from positive and negative information when the positive and the negative are equally informative. While the results of the first experiment are consistent with previous findings, they are not supported by the results of the second experiment. The inconsistency was hardly expected.

References

Bruner, Jerome S., Goodnow, Jacqueline J., and Austin, George A. A Study of Thinking. New York: John Wiley & Sons, Inc. 1956.

Hovland, Carl I., and Weiss, Walter. Transmismission of information concerning concepts through positive and negative instances. J. exp. Psychol., March 1953, 45, 175-182.

CHAPTER VIII

REINFORCEMENT OF ORIGINALITY OF RESPONSE

Introduction

During the present decade there has been a revival of interest in the function of education in developing creative talent, an interest which had reached a peak a quarter of a century earlier under the sponsorship of the progressive education movement. The typical position taken is that creative abilities are trainable in much the same way as are other abilities and that the exercise of these abilities coupled with appropriate rewards will promote development. Some have even proposed that individuals can learn specific techniques which will improve their creative performance. Such proposals are attractive to the promoters and their followers but they lack any kind of theoretical justification. The purpose of this paper is to present a theoretical position regarding the nature of creative abilities and to test empirically some of the implications of this theory for training.

The study was concerned with determining the trainability of children in the capacity to produce original responses. The general procedure was that of measuring the ability to produce original responses before and after the training procedure and then comparing the scores on the two measures. The assumption underlying the present study is that the production of an original response to a task such as the Guilford *Unusual Uses Test* depends upon the previous establishment of an association between the element of the task that constitutes the basic stimulus and the unusual use that is given as a response. The assumption is that the individual does not have the capacity to generate such an association in the testing situation itself, but only that he can utilize the associations already established. While he may have much greater access

to such weak, but previously-established, associations under some conditions than under others, one would not expect that that training would have any effect on the ability to produce original responses in the task situation under consideration, for the repertoire of such associations is strictly a function of previous training. Furthermore, accessability to such associations would not be expected to be amenable to training since accessability is postulated to be a function of the degree to which environmental influences inhibit or facilitate the production of such responses. Nevertheless, there is still the possibility that any training effects that are found may be primarily motivational, that is to say, training may not increase the number of associations available and neither may it increase the person's access to these associations but it may result in him wanting to produce (not generate) more remote associations (i.e., original responses). In such a case an experimenter, who during training becomes a source of motivation for producing original responses, may continue to be such a source of motivation when he administers the criterion tasks after training. The motivation theory appears to be the most plausible one at the present time. The studies covered in the review of the literature that follows give some support to this position.

REVIEW OF RELATED RESEARCH

Studies of Attempts to Train Original Behavior in Laboratory Situations

Maltzman (1958, et al.) was interested in the conditions that facilitate the production of uncommon or original responses. These

responses are assumed to be lower in the response hierarchy and to have a weaker excitatory potential than common responses. The procedure employed in his first study was an attempt to arrange a situation in which subjects would be induced to give responses low in the response probability hierarchy. The study further experimented with verbal reinforcement of uncommon responses, and instructions to be original during training. Subjects (N=120), university psychology students, were all orally presented with an initial stimulus word list for training. They were given the usual free association instructions to respond only once as quickly as possible with the first word that came to mind. They were then divided into three groups. The control group was then given a test list. Two experimental groups were presented again with the training word list four more times and instructed to give a different response each time from the ones given previously. At the end of the training period the control groups were then given the test list. One of these experimental groups was administered a verbal partial reinforcement schedule of "good" by the experimenter for approximately every fifth uncommon response. The Guilford Unusual Uses Test was administered to all subjects after the last repetition of the training list. Before the administration of the Unusual Uses Test half of the subjects in the three groups received special instructions to be as original as possible and half were not. Results with the test list confirmed the hypothesis that repeated presentations of the initial list with instructions to give different responses each time would increase the uncommonness of test responses. Almost identical results were obtained for the two experimental groups (with and without reinforcement), indicating that the verbal reinforcement employed did not influence uncommonness

of responses. The reinforced experimental group gave significantly more original responses than the control group (P≐.02 level of confidence) while the difference between nonreinforced experimental and control groups was in the same direction but not at an acceptable level of confidence. Significant differences were found on the Unusual Uses Test in favor of experimental subjects receiving instructions to be original on the test (P=.01 level of confidence).

The final conclusions drawn from the study indicate that originality of word associations can be facilitated by practice in free association as well as by instructions. Verbal reinforcement of uncommon associations did not produce significant increases in originality. Maltzman stated that: "It is possible that a greater number of reinforcements and/or a different reinforcement schedule would produce a more pronounced effect than the one obtained...These effects, however, require further confirmation before they can be interpreted unambiguously."

An additional problem concerning reinforcement presented itself within the context of the foregoing experiment. In this latter study an experimental group had been verbally reinforced for unusual responses during the training phase, but this procedure had failed to produce an effect on the test series. On the other hand, the increase in the production of original responses is indicative that some sort of reinforcement was possibly operating, although not under experimental control. It might have been that the verbal reinforcements used simply failed to contribute to the reinforcements that were already operating in the situation.

Maltzman (1960, <u>et al</u>.) has reported a series of several additional laboratory studies indicating positive results in attempts to increase the production of original responses or to increase the originality of responses. In the first study of this series, university psychology students used as subjects (N=292) were administered a training list of twenty-five words taken from the Thorndike-Lorge count (1944) to which they were to free associate. They were all given the usual free association instructions to respond as quickly as possible with the first word that came to mind. Following this procedure, the subjects were divided into two control and three experimental groups. One control group was given a test list of twenty-five new stimulus words with the same instructions as given for the training list. A second control group was given five additional presentations of the same training list and instructed each time to try and give the same response to a given stimulus word each time. One experimental group was given five repetitions of the new test list and instructed to try and give a different response to a given stimulus word each time. Two other experimental groups were each given different test lists of 125 different words following the initial training list with the same instructions to free associate and respond with the first word that came to mind. One group received a list of words with a low frequency count, from the Throndike-Lorge count of words occurring not more than six times per 1,000,000. The other received words with a high count; that is, from among the 500 most frequently occurring words in the count. All subjects were given the Guilford <u>Unusual Uses Test</u> immediately after the completion of the free association test lists. Results indicated that the experimental

group given instructions to respond differently to each repetition of each word in the test list was significantly more original than the control groups (P=.001 level of significance). A simple analysis of variance of the original responses obtained from the Unusual Uses Test showed that the experimental groups receiving five repetitions of the test list with instructions to give a different response each time gave significantly more unique responses than any of the other four groups (P=.001 level of significance). None of the groups differed significantly in terms of their fluency of common responses on the Unusual Uses Test. A basic assumption in the use of the training procedure employed with the one experimental group was that instructions to be different and original each time when responding to repeated stimulus words would increase the uncommonness of responses. Results showed that this training method had the expected effect.

The second experiment used university psychology students (N=251) who were again presented an initial training list and final test list of stimulus words with the free association instructions as before. The stimulus word list used for training and the subsequent test were the same as those of the last experiments; however, the two lists were reversed so that what was previously the training list was now the test list. Guilford's Unusual Uses Test was administered after the completion of the free association test list. The control group received the single training list followed by the test list. Four experimental groups received different treatments as follows: experimental group I received five repetitions of the training list with instructions to give a different response for each stimulus word each time; experimental group

II was asked to respond differently to five repeated presentations of six items from the Unusual Uses Test; a third experimental group was presented with 125 pairs of words and was instructed to underline the number of the pair matching more readily the stimulus word presented orally from the training list; and a final experimental group received the same response pairs as group III but without the stimulus word but was instructed to underline the member of each they thought to be more familiar. Results obtained in this experiment confirm the results of the first experiment in that experimental group I differed significantly from each of the other groups (P=.001 level) on mean originality scores on the free association test. The Unusual Uses Test was administered to all groups after the completion of the training procedure. Repeated evocation of different responses to free association stimulus is followed by an increase of originality of responses to new stimulus words and to items on the Unusual Uses Test as compared to a group not receiving such training. However, training with material similar to that provided by the Unusual Uses Test produced no effect.

A third experiment by Maltzman (1960) was conducted to determine the effect of amount of training on a free association word list. Amount of training can be varied by either increasing the number of items with which to free associate, or increasing the number of repetitions of the same items presented. Both variations were employed in this experiment. Subjects, (162) introductory psychology students, were given the same initial training and final test consisting of a free association word list as in experiment II. All experimental groups received single presentations of the training and test lists. However, prior to

receiving the criterion free association test list four groups were exposed to different experimental treatments. One group received ten repetitions of six items from the Unusual Uses Test while a second group received twenty repetitions of the same items. Two more groups received six additional items from the Unusual Uses Test repeated either five or ten times. Thus, treatments consisted of either ten or twenty repetitions of six Unusual Uses Test items or five or ten repetitions of twelve similar items. No evidence of a facilitating effect of training by repeated evocation of different responses to the same stimuli was obtained. It was likewise found that no significant differences occurred with varying the number of unusual uses items on the training list prior to the free association test for originality. There was no significant change in uncommonness of responses on the free association test. Maltzman et al. concluded that repeated presentations of unusual uses items were ineffective because they did not produce large enough increase in uncommonness of responses evoked during training. Therefore, a fourth experiment was conducted using the free association situation for training since reliable effects upon originality had previously been obtained under these conditions.

Experiment IV of the Maltzman (1960) series again utilized (N=201) introductory psychology students. Two new stimulus word lists used for initial training and free association testing were drawn from Woodworth and Wells (1911) on the basis of their tendency to evoke opposites. There were twenty-four words in the training list and ten words in the test list.

The Guilford Unusual Uses Test was administered after the test list. The control group received the test list of 10 new stimulus words following completion of the initial twenty-four word training list. Three different experimental groups received one, five, and ten repetitions of the training list prior to the presentation of the test list. Results indicated no significant differences between experimental groups. Results from the free association test of originality indicated that each of the three experimental groups gave more original responses than the control group (P-.001). The results of *t* tests indicated that the three experimental groups did not differ significantly from each other on the Unusual Uses Test. Each of the three experimental groups gave significantly more unique response on the Unusual Uses Test than the control group (P=.05 for the five repetitions group versus control, and P=.001 for the other two experimental groups versus control group). A fifth experiment was conducted by Maltzman (1960) in which he hypothesized that if originality can be learned, according to the principles of instrumental conditioning, then it should show some degree of persistence. Subjects (N=77) university psychology students, were administered stimulus word lists from the Thorndike-Lorge count as in studies I, II and III previously reviewed. In addition to the initial free association training list, a test list and the Guilford Unusual Uses Test were administered. All subjects received the training and test lists. Two control groups received the free association test and Unusual Uses Test I 48 hours following training. Two experimental groups experienced comparable delays but, as in the previous experiments received training by receiving five repetitions of the training list with instructions each time to

respond differently to each stimulus word and to be original. Significant differences were obtained for the originality training effects of the experimental groups as compared with the control groups (P=.001 level) on the free association test following training. Significant training effects after delays from one hour to two days on the free association test of originality were obtained (P=.05 level of significance). There was no significant decrease in the number of unique responses when the delay was increased from one hour to two days.

As a final conclusion to the series of five experiments reviewed herein, Maltzman has stated: "The results of the experiments lend some support to the hypothesis that originality is learned behavior and varies as a function of the same antecedent conditions as other forms of operant behavior. Effects of originality training may persist for as long as two days...originality, at least on the free association test, also varies as a function of the number of repetitions of the training list."

A few comments must be made on certain striking features of the Maltzman series of experiments. First, there is a certain oddness about the finding that free association activities result in an increment in score on the <u>Unusual Uses Test</u>, but that training on unusual uses items does not. Certainly one would <u>not</u> expect a greater amount of transfer to occur from a task very different from the criterion task than from a closely similar task. Such a finding would be unique in the entire literature of learning research. While Maltzman has an explanation to offer for this paradox, an alternative and much simpler explanation is that he was not dealing with a learning phenomenon at all, but rather his results should be interpreted as representing a facilitation phenomenon

similar to that manifested by a pitcher when he warms up in the bull pen.

A second important point to note is that, throughout the Maltzman series of experiments, the experimenter was both the administrator of the training task and the administrator of the pre-test and training task. Under such conditions, the training tasks may help to define what the experimenter is requiring of the subject and, having learned what is required of him, the subject may show an improved performance on the criterion task. Under such conditions, the presence of the experimenter cues the performance of the subject on the criterion task. One suspects that if the criterion task were given by a different experimenter and if the criterion task were not connected in any way with the training task that the so-called training effect reported might well vanish.

The puzzling nature of the Maltzman findings are rendered even more puzzling by the fact that Anderson and Anderson (1963) trained subjects in a situation requiring them to name as many uses as they could of objects and found that the training transferred to a <u>Novel Uses Test</u>. The fact that Maltzman found no effect of such training, while Anderson and Anderson (1963) did, can perhaps be attributed to the extended nature of the training in the latter study and the very limited training given in the former. It should be noted that Anderson and Anderson, like Maltzman, also used the same experimenter to run the training sessions and to administer the criterion tests, so the effect reported may well not be a genuine learning phenomenon.

Studies of Attempts to Train Original Behavior in Academic Situations

Parnes and Meadow (1959) investigated training effects of a 30-hour course in creative thinking using the so-called "brainstorming" technique outlined by Osborn (1957). The usual description of the technique is that of withholding or deferring judgment while producing ideas. Parnes writes that: "The basic thesis of this technique (brainstorming) is that creativity is encouraged by the temporal segregation of hypothesis formation and the judicial evaluation of the adequacy of hypotheses." The training course emphasized the brainstorming technique throughout with pre- and post-training devices measuring the variables of quantity and quality of ideas and three personality variables of need achievement, dominance, and self-control. Subjects (N=162), university school of business day and night students, were equally divided in an experimental group and two control groups. Eleven measures comprised the pre and post assessment devices which were divided between the two control groups. One control group was administered six creative ability measures including the Guilford battery and the _AC Test of Creativity_ while the other was administered five personality measures including the _Thematic Apperception Test_ and the _California Psychology Inventory_. Control subjects were matched with experimental subjects on age, sex, and the _Wechsler Vocabulary Score._ The experimentally trained group was administered all eleven measures by their course instructor at the beginning and end of the semester's course. Results comparing the experimental and control groups indicated significant differences on both quantitative and qualtative measures of five creative ability tests in favor of the

experimental subjects (P=.01 level of significance). Parnes and Meadow stated that: "Since the instructor carefully avoided practice on any objects even remotely similar to the type of objects which appeared on the tests, the results do indicate generalization of this training." However, the trained student did receive instruction and practice on types of problems similar to those included in the tests. Of the three personality measures used, the CPI Dominance scale indicated a significant difference, or an increase in Dominance, of the experimental groups as compared to the control group (P=.05 level). The other two personality variables measured did not yield significant differences. The final conclusion of the study was that training in the technique of brainstorming produced a significant increment on measures of quantity and quality of ideas and on the personality variables of dominance. It is interesting to note that the same instructor initiated all tests as well as conducted the training program for the experimental group. An interesting question arises whether the difference in creative performance on the tests was attributable to practice or to an artifact in which subjects were responding to the tests the way they thought the instructor wanted them.to.

An experiment designed by Parnes and Meadow (1960) to evaluate the persistence of the effects of a course in creative problem solving emphasized the brainstorming technique. An experimental group that had taken a creative problem solving course eight months previously was compared with a control group of students who had just enrolled in a similar course. Experimental and control groups (N=48) were matched for vocabulary ability on the basis of the <u>Wechsler Adult Intelligence Vocabulary Test.</u> The

experimental subjects had previously taken the course in creative problem solving at least eight months prior to the study and the control subjects were registered for the course but had received no training. Both groups were compared on six measures of creative ability which included the Guilford Apparatus Test, Unusual Uses Test, Plot Titles Test, and Possibilities Tests and quantity and quality measures on the Uses Test from the AC Test of Creative Ability. No subject in either group had ever before taken these creative thinking tests. Results indicated that the experimental subjects trained eight months before outperformed the control group. Four of the six differences between the groups were statistically significant (Beyond the P=.05 level of significance). Differences between the experimental and control subjects for the Apparatus and Unusual Uses measures did not meet the .05 level of significance. The authors stated that: "Results indicate that increased productivity in creative thinking produced by the creative problem solving course persists for a period of eight months or more after completion of such training."

These experimental studies have attempted to control and manipulate variables in a deliberate development of fluent and original thinking as factors of creativity. Only a few studies have been found that were designed to develop creative thinking in school classroom situations. Two such studies will be discussed because of their relevance to the present investigation. Hutchinson (1963) conducted four fifty-minute in-service teacher training seminars in which he instructed classroom teachers in processes of divergent thinking. Likewise, students of these teachers were instructed in the group methods of brainstorming. Subjects, (N=256) seventh grade students, were divided into two groups, four experimental

and four control classes matched by mental age and sex. The four control classes were taught by teachers who did not participate in the training seminars and instructed their classrooms in the usual manner.. The four experimental classes were taught by in-service trained teachers who were shown how to modify their teaching methods to include classroom work involving divergent and evaluative thinking. Ten measures of creative thinking and also subject-matter tests were administered towards the beginning and towards the end of the experiment to both groups. Both groups of classes were taught a two-week social studies unit on transportation and communication. Tape recordings were made of all class sessions of both groups and classified according to the Aschner-Gallagher Classification System. There was no significant difference between the two groups in the amount of gain shown on subject matter tests, but the experimental group had significantly greater gains in four of the ten measures of creative thinking (three at the P=.05 level of significance, one at the P=.01 level). These might be attributed to training though other explanations are possible. The author stated that by teacher training and modification of instruction methods for experimental classes, there was a distinct change in verbal response categories during classroom instruction. The experimental group decreased 30 percent in the routine response category, increased 20 percent in the cognitive-memory category, and increased 23 percent in the divergent-thinking and evaluation category.

Knight (1963) also designed a study to measure the training effects on fluency, flexibility, and originality on a mixed grade of fifth and

sixth grade school children (N=28). A control group in another school was matched with the experimentally trained group on age, intelligence, sex, general scholastic ability, and occupational level of the family wage earner.

Seven tasks were administered to both the experimental and the control groups both before and after the training program and from these tests fifteen scores were derived which could be combined into measures of fluency, flexibility and originality. Training consisted of exercises involving psychological operations closely similar to those involved in the evaluation devices. Some of these training tasks were performed by pupils working alone but some required group activity. The training program extended over the school year. Knight was the teacher of the group exposed to the special training and was also the person who administered the tests at the beginning and end of the school year. Significantly greater gains were found for the experimental group over the control group on 14 of the 15 measures of creativity at levels of significance which ranged from .05 to .001. Achievement tests also given at the start and finish of the school year indicated that any gains made on the creativity tests were not made at the expense of the commoner forms of the school achievement.

The studies of attempts to train original responses leave the reviewer with considerable discomfort concerning the interpretation to be given to them. First, there is the disconcerting fact that, without exception, the same experimenter served as trainer and as test administrater. The possibility remains that training may have taught the subjects

to produce a certain kind of response for the experimenter. The subjects have learned not so much to be creative but to do what the trainer expected of them. The results of this might be that when they came to the final criterion test they knew better what was expected of them than they had when they took the initial test. In a sense, the effect of training may have been that of giving subjects a better understanding of what was expected of them when faced with a test of creative ability. Much the same effect as that produced by training should be produced by using suitable directions on the test. This hypothesis can be examined since a number of studies have now been conducted which investigate the effect of administrative directions on performance on such tests. These studies must now be examined.

Studies of the Effect of Administrative Directions on Performance of Tasks Requiring Creative Behavior

Christensen *et al.*, (1957) experimented with fluency and originality on tasks emphasizing inventiveness. It was hypothesized in this study that production of original responses on an open-end test calling for creative thinking would be relatively constant across time. It was further hypothesized that more original responses would come later in the production period, and that with instructions to be clever the total quality of responses would be reduced while more clever responses would appear. Each of these hypothesis were treated in separate parts of the study. Criterion performance measures consisted of the *Plot Titles Test*, an impossibilities test, and the *Guilford Brick Uses Test*. Subjects (N=400) were Air Force cadets and college students. The relation of

number of responses to time for the written creative thinking tests used showed less decrement with time as compared to verbatim-recall tasks. Uncommon responses increased in frequency with time. The relation of remoteness of responses to temporal order of occurrence showed significantly greater numbers in the second-half over the first-half of the test performance (p=.01 level of confidence). The increase of cleverness of responses with time did not occur as predicted. When two plot titles tasks were administered to two groups with instructions to one group to give only appropriate titles, and to the other to give appropriate, clever and novel titles the following results were obtained: A significantly greater number of clever responses were produced by instructions to give clever and novel titles (p=.01 level) even though there was a smaller average total number of responses produced under this condition. As expected, fewer total responses were produced when subjects were instructed to be clever. When these were compared to a group of subjects receiving no special instruction to be clever, the correlation between number of clever responses under instructions to be clever and no instructions was .62. Christensen stated: "It appears that individuals are more or less disposed toward making clever responses whether or not they are explicitly instructed to do so." Other studies have shown a decrease in quantity of total output under instructions to be original or clever. This effect is probably due to self-imposed censorship of the subject which facilitates rigid limits of rejection for so-called non-clever responses.

Meadow, Parnes and Reese (1959) designed another study to measure subjects' ability to solve a problem requiring creative ability either

under instructions to express all the solutions they could without evaluation (brainstorming) or with instructions to express only solutions of good quality and which involved a penalty for solutions of bad quality. The penalty was the subtraction of ideas rated "bad" from the total, and this information was given to the subjects. The quality of the solutions was later judged by a trained rater. Subjects were (N=32) college students who had been enrolled in a course in creative problem solving which stressed the brainstorming procedure for problem solutions. Subjects were randomly divided into four experimental groups. The problems were taken from the <u>AC Test of Creative Ability</u>. One fourth of the subjects was given one of the two problems either under brainstorming or nonbrainstorming instructions. Each response was rated for uniqueness and for value on a 3-point scale. A response was designated as "good" if the combined uniqueness and value score was above five. The number of good responses was then expressed as a percentage of the total number of responses which constituted each subject's creativity score. Results showed that significantly more good solutions were produced under the brainstorming instructions than under nonbrainstorming instructions (P= .01 level of significance). Significantly more good solutions were produced when brainstorming instructions were given in the first test period than when given in the second test period following nonbrainstorming instructions ($.05 < P > .02$). One main point should be emphasized with respect to interpretation of the study; all subjects were initially trained in the brainstorming technique. Because of this factor another study was devised using nontrained subjects.

Parnes and Meadow (1959) designed this second study to determine the effectiveness of brainstorming instructions with untrained subjects and to compare the data thus derived with the data from the subjects previously trained in brainstorming. University undergraduate students (N=52) were used. Grade point average was used for matching formerly trained subjects to untrained subjects. Trained subjects differed from untrained subjects in having been exposed to a one-semester course in creative problem solving utilizing the brainstorming principle. The same two problems from the AC Test of Creative Ability were used to provide the criterion scores as in the previous study. The untrained subjects were given one problem with the second following immediately thereafter; each problem was administered either under brainstorming instructions or nonbrainstorming instructions. The same procedure was followed for trained subjects. The data showed that more good ideas were produced under the brainstorming instructions than under the nonbrainstorming instructions, and this effect was greater in the second test period than in the first. Subjects formerly trained in a course emphasizing the brainstorming technique produced a significantly greater number of good quality ideas when using the technique than did untrained subjects (P=.01 level of significance). An additional result reported a positive correlation between quantity and quality of ideas either under brainstorming or nonbrainstorming instructions (r=.63 to .81). The point to be made here is that with either specific instructions to be fluent or original under brainstorming or nonbrainstorming instructions some subjects were able to produce original ideas.

Another study by Parnes (1961) involved two experiments designed to test the hypothesis that extended effort in idea production during a single session will lead to an increasing proportion of good ideas. Both experiments used university undergraduate students divided into two groups. Group one (N=146) comprised untrained students who were presented the Hanger problem from the _AC Test of Creative Ability_. Instructions were given these subjects to think of all possible uses of an ordinary coat hanger in the time limit of 5 minutes. Scoring yielded a quality score defined as uniqueness and unconventionality; and unduplicated responses were counted for a quantity score. Interrater reliability for ratings of fifty subjects selected at random was .74. Each subject's data was divided into halves on the basis of total number of ideas and an analysis of variance was used to determine differences in quality of ideas between the first half of each subject's ideas and the second half. Results indicated highly significant differences (p=.001) between halves in favor of the second half. A second group (N=42) of students that were enrolled in a creative problem solving course which consisted of training and practice in brainstorming was tested. Near the completion of the course they were presented the Hanger problem as in experiment 1 above with treatment time extended to 15 minutes, and it was again scored for quantity and uniqueness of responses. The mean for the last third was significantly greater than the means for both the first third (P=.01) and middle third (P=.05) of the list. The means for the first and middle thirds did not differ significantly.

The data that has been presented up to this point indicates that the directions given in administering a test of creativity, and the

interpretations given to them by the subjects are important factors in determining the general level of the scores. This effect is so substantial the possibility exists that the "training effects" reported by Parnes and others may well be a result of a changed orientation towards the criterion test produced by training rather than a genuine learning effect. In this connection a study by Gerlach, Schutz, Baker, and Mazer (1964) throws some light on the matter. In this study, two tests of creativity were administered with differing administrative directions. Altogether six different sets of directions were used which included the standard directions, brainstorming directions and what were referred to as "criteria cued" directions. The latter directions spelled out in considerable detail the performance required on the test. This set of directions yielded higher scores on the test than a trained group previously reported by Parnes (1961). The students in the Parnes experiment who had been trained in brainstorming and who were given the test with brainstorming directions had a slightly lower mean than the untrained group in the Gerlach, Schutz, Baker, and Mazer study. The data suggests that the orientation of the subjects towards the criterion task may be the crucial factor operating in most of the experimental studies of training in creativity that have been reviewed in this chapter including those of Maltzman. Another very important finding to note is that, when very complete directions were given, the subjects showed no tendency to produce the more original responses in the later part of the testing time. This finding negates the notion that practice during a single session produces some mysterious learning effect. It appears to be the task orientation which determines, to an important degree, the order in which responses will be evoked.

<u>An Overview of Studies on the Training of Creative Talent</u>

The review of the literature which has been undertaken reveals an odd inconsistency between the state of knowledge in the field and current educational practice. A brief inspection of the course offerings of institutions concerned with the training of teachers reveals that courses on the development of creative talent in children are widely offered as though there were an extensive and well-developed body of knowledge to transmit. Our review of the literature shows that the area is one which has stimulated experimentation and exploration, but that the outcome of most research is a set of controversial conclusions. Those who have engaged in research in this area cannot be blamed for the lack of definitive findings. Attacks on new areas at the frontiers of knowledge typically produce what appear to be inconsistent findings until experimental techniques are evolved capable of producing stable results. One is reminded of the fact that research on learning has had to have taken place over almost an entire century before laboratory techniques emerged which yielded highly consistent results. Only in the last fifteen years has knowledge of learning phenomena developed to the point where laboratory demonstrations can be givin in front of entire classes of students and with assurance that learning will occur exactly as predicted. Such control over behavioral phenomena is evidence of a mature science. The area of research on creativity is at the beginning of its history and the knowledge which such research has produced must be considered to be highly uncertain. One should, perhaps, regard it as knowledge suitable for stimulating the future work of scientists rather than

research which has direct applications to problems of education. These introductory comments serve to suggest that the study which follows is not intended to provide any immediate solution to an important educational problem, but serves the purpose of helping to determine the nature of the task of developing creative talent.

THE PROBLEM

The purpose of the present study is to provide relatively prolonged training in the making of original responses and then to evaluate the extent to which the training is effective. A major difference between the present study and previous studies is that care has been taken in the present study to eliminate from the evaluation procedures certain conditions which have, in previous studies, appeared to produce spurious effects. Specifically, the use of the same experimenter both for the presentation of the training materials and for the administration of the evaluation tasks has been replaced by a procedure in which the experimenter administered the training tasks, but the teacher administered the evaluation instruments as a part of the regular classroom routine. Thus any spurious effect which the experimenter could have on the procedure is eliminated. In addition, fairly prolonged training is given involving in the one case the practice of original responses and in the other practice together with reinforcement. The expectation was that the training effects which have appeared in previous studies would vanish under the better controlled conditions of this study.

METHODS AND PROCEDURES

This chapter presents the experimental procedures of the study in terms of: a) subjects; b) training tasks -- their characteristics, administration and scoring procedures; c) criterion tasks -- their characteristics, administration and scoring; d) training conditions of reinforcement versus nonreinforcement; and 3) the design.

Subjects

Those who participated as subjects in the experiment were 209 elementary school students (110 girls and 99 boys). All Ss were sixth grade students from three elementary schools. Two of the schools are located in Bountiful, Utah, a suburban residential area north of Salt Lake City. Three sixth grade classes in one of these schools were trained for fourteen days by an experimenter on a task involving creative abilities and were reinforced by a partial reinforcement schedule each day of training. These Ss are referred to as the reinforced experimental trained group and are designated by their respective school room numbers of 10, 11 and 12. Three more sixth grade classes at the other school were likewise trained for the same fourteen days but were not reinforced during training. These Ss are referred to as the nonreinforced experimental trained group designated by their respective school room numbers of 22, 23, and 24. Criterion tasks were administered during, after and two weeks following training by the regular classroom teacher in all six classrooms of both experimental trained groups. A third school, located in an old residential section of downtown Salt Lake City, was used as the control group for certain purposes. Three sixth grade classes at the

latter school were administered only the criterion tasks by an experimenter over the same time interval as the experimental trained groups. This group of Ss (N=47) served as a control group not receiving training and is designated as control group classes 1, 2 and 3.

Neither of the two experimental schools grouped children in a grade with respect to intelligence or achievement. The range of intelligence measures for this sample of elementary school students as subjects was from 81 to 146. The three classes of the control group were grouped by intelligence scores into two classes: a gifted class (room 3) made up of students with intelligence test scores over 110, and the other two rooms (1 and 2) made up of the remaining students. Table 8.01 indicates the mean and standard deviations of the intelligence test scores by class measured by the California Mental Maturity Test which had been administered to all Ss in the fifth grade. The table reports data for only those pupils used in the analysis of the data. The basis for this selection will be described in later sections.

At the completion of the study, test data on all Ss who had been absent on any one of the counterbalanced training days or on any one of the criterion testing days was eliminated. The remaining Ss were matched as nearly as possible across reinforced and non-reinforced groups by intelligence test scores. Complete test data on a total of 96 Ss who were matched by intelligence scores and had been present throughout the complete five week training and testing program then remained as the experimental sample for analysis.

TABLE 8.01

INTELLIGENCE TEST SCORE MEANS AND STANDARD DEVIATIONS FOR EXPERIMENTAL AND CONTROL GROUP CLASSES FOR CASES USED IN THE STUDY

	N	Mean	Standard Deviation
Reinforced Experimental			
Trained Group*			
Rooms 10	16	114.25	15.04
11	17	121.11	11.44
12	15	113.87	12.38
Nonreinforced Experimental			
Trained Group*			
Rooms 22	17	120.70	12.18
23	13	113.15	14.36
24	18	114.94	12.96
Control Group			
Rooms 11	15	91.90	10.3
2	18	90.53	15.1
3	14	118.64	9.9

*California Mental Maturity Test
**Pinter Verbal Intelligence Test

Training Tasks

Task Characteristics

The training task had a form similar to the Guilford Uses Test (1962). It consisted of a list of ten names of common things (objects) presented in print form on paper. This task is referred to as the training word lists. Fourteen training word lists, one for each day of training, were prepared on multilith forms for the study. The forms were prepared in a manner so as to differentiate them from the "ditto" form of the teacher controlled criteria tasks. These training word lists presented by the experimenter were purposely made to appear distinct from a classroom exercise. Even though fourteen different word lists were used, three of these were counterbalanced across the three classes of both the reinforced and nonreinforced experimentally trained groups on the 7th, 11th, and 14th days of training. Printed instructions on each word list directed the Ss to find a different use for each thing, to be original and creative, and to use their imagination.

Administrative Procedures

One list of ten names of common things which comprised the training word list was presented each day for the fourteen training days by an experimenter following a short orientation procedure. Subjects were instructed to think of one unusual use for each object. The following training instructions were given each day by the experimenter before administering the task:

> This is an exercise to see how well you can use your imagination for thinking up unusual uses for common things which are all well known to you. Try to think up a different use for each thing than the one you think your neighbor would put down. Be original, do not suggest a use for the thing which most people would think of. You are to think up <u>one</u> unusual use for each thing and write it down as soon as it

> comes to your mind. Work rapidly, but try to think up a use for the thing in a way that it was not originally intended to be used. If you cannot think of a different use for the thing after a short time, skip it and go on to the next thing. Be brief, do not worry about your spelling or complete sentences. Just write the first unusual use for the thing that comes to your mind.

The orientation procedure preceded the administration of each training word list and was standardized across all six experimental classes at the two schools. The experimenter read the training instructions aloud each day and then picked some item at random in the room; such as a blackboard eraser, or a pencil, etc., to use as an example. He would then ask the class for their ideas on how they might use the object differently while pointing out its inherent attributes or qualities, and informing Ss to think of using its different parts or the whole object in some new or unusual way. An example of the typical training orientation used by the experimenter is as follows:

> Try to think of a different use for this thing than the one you think your neighbor would put down. Be original, do not suggest a use for the thing which most people would think of. Some things have their common use within the name of the thing such as tooth brush. The common use would be to brush your teeth, but try to think of a new use for a tooth brush such as to apply paint in cracks. You may use all or any part of the thing, or change its form. For the tooth brush you may think of removing the bristles and using its plastic handle as a tool for working with clay. Other things do not have their use implied within their name, such as a brick. Here you can be as original as you want. How would you use it differently? Here are some examples
>
> Number 1. brick ----------- powder it and make red paint
> Number 2. newspaper ------- to make a kidnap note
> Number 3. paper clip ------ as a fish hook
>
> You are to think up only one unusual use for each thing and write it down as soon as it comes to your mind. Work rapidly but try to think of a use for the thing in a way that it was not originally intended to be used. If you cannot

think of a different use for the thing, skip it and go on to the next one. Be brief, do not worry about your spelling, just write the first unusual use for the thing that comes to your mind.

Are there any questions?

Each class was allowed to run-off ten to twelve ideas orally before proceeding on to the written training task for the day. The training procedure and time for handing out papers took about 8 minutes for each class. A strict time limit of 5 minutes was imposed for each training word list. At the end of this 5 minute period, time was called, all the papers were collected, and the experimenter moved on into the next classroom.

Reinforcement

A verbal reinforcement schedule was used for the three classes that constituted the reinforced experimental trained group. At the first of every training period, following the first day, the training word list papers of the preceding day were returned to each S. These had been scored and some of the responses were reinforced with "excellent," "very good," or "good." These words were stamped in red alongside those ideas judged for their originality. Since no scoring guide was available for the training task, the basis for choosing responses to be reinforced was the infrequency of the responses among the population sampled. One of the members of the Bureau of Educational Research staff daily reinforced ten percent of those responses determined to be original in terms of infrequency. The most infrequent, relevant responses in the population sample were reinforced with an "excellent." The next most infrequent relevant responses were reinforced with a "very good," and the third most

infrequent responses with "good."

In this manner, each S of the three reinforced classes was given daily knowledge of results on how original his or her ideas were the day before. The pupils responded enthusiastically to the daily return of their papers. These papers were first distributed and then collected before the training word list for the day was handed out. The initial training procedure was then carried on by the experimenter. The Ss of the non-reinforced classes were given no knowledge of results on their daily training performance since their papers were not returned each day. All teachers remained in each classroom during every training period either observing the procedure in the back of the room or working at their desks.

Scoring

The responses to the training word lists for the experimental subjects were scored for originality. The forms containing the answers were coded by number and all names and any reinforcing statements stamped on them were removed. This was done to eliminate class or treatment identity to prevent contamination of the results.

Training task word lists for the first day of training were distributed evenly among eight members of the Bureau of Educational Research staff for practice scoring along with a set of scoring instructions. The scoring procedure consisted of three steps. First, scorers were to eliminate all irrelevant responses, or those that were irrelevant. Of the remaining responses, scorers were asked to choose ten percent of the most infrequent and unusual uses for the items listed. From the ten percent chosen as original responses, scorers were

then asked to select one third of these as the most unusual or creative responses. The procedure for scoring a response as original was adapted from Torrance (1962) and was based upon the infrequency of the response in the sample.

After the papers from the first day of training had been scored once, they were collected and again redistributed for a second scoring among the same eight scorers. In this manner all Ss training papers for the first day of training were scored twice for originality. Interscorer contingency coefficients between scorers for individual items ranged from .11 to .36 which were all significant at the .001 level of confidence except one at the P=.05 level. The over-all contingency coefficient for all data combined was 0.22.

After this practice scoring procedure the eight scorers set about the task of scoring the counterbalanced training word lists administered on the 7th, 11th, and 14th training days. The procedure was adopted of coding the papers, shuffling, and then distributing them among the eight scorers. After these had been scored once for originality, scoring sheets were removed and replaced with new ones after the papers had been reshuffled. They were again distributed for a second scoring.

Criterion Tasks

Task Characteristics

Teacher administered criterion tasks were used for evaluating transfer from the experimenter training exercise to the classroom teacher situation. Two types of criterion tasks were utilized. 1) The first was a written task called the Ask-and-Guess Test (Torrance, 1958) - a paper-

and-pencil test containing three parts, each requiring the student to respond differently while viewing picture prints from Mother Goose stories. Three 35mm colored slides were projected on a screen, one for each test period, for group, classroom testing. The first part of the test requires the student to ask questions about the picture which cannot be answered directly from the picture. In the second part the student is asked to make guesses or formulate hypotheses about the possible causes of what is shown in the picture. The third part asks for consequences, both immediate and remote, of what is happening in the picture. The three Mother Goose picture prints used for this Task are identified as P_1, P_2, and P_3.

2) The second set of teacher-controlled criterion tasks consisted of three written exercises designed for the study. These are referred to as the A, B, C, criterion tasks: The first, Task A-Lost, asks the student to list unusual ways he would use the things mentioned in the problem in order to feed and protect himself in a situation involving being lost in wild country. He is asked to think of ingenious uses for five things listed in the problem; such as, safety pin, ball of string, leather belt, etc. This task is similar to the unusual uses training task but is stated as an open-ended problem situation. The second, Task B-Rope Problem, is a modification of the Maier pendulum problem. The student is asked to think of different uses for a given number of things in order to solve the problem of tieing two ropes together that are not long enough for both to be reached. The third Task C-Rocket Ship Situation, asks the student to list consequences of things that would happen as a result of a given situation.

All criterion tasks were reproduced in the same way as other school

materials-by ditto machine on a cheap white paper. The purpose of this was to lead the pupils to believe that they were undertaking a school assignment. They were administered by the regular classroom teacher (except in the case of the control group) and the teachers were asked to make no reference to the experiment.

Administrative Procedures

Both types of criterion tasks were administered before and after training and two weeks following training by each classroom teacher of the six experimental classes involved in the study. These tasks were administered by the experimenter to the three control group classes over the same time interval. The following complete instructions were used by each classroom teacher and the experimenter for administering the three parts of the *Ask and Guess Test*.

General Directions (Flash picture on screen)

> There are many ways of finding out the things you want to know and need to know -- the things you need to know in order to evaluate a situation and plan your strategies for coping with the situation. Our favorite ways of finding out things is by asking questions, observing, and making guesses. To the extent that we can, we check these guesses against other things that we can observe or know from other sources. The three tasks in this booklet are based on a picture which you will be shown shortly. You will be able to look at the picture throughout the three tasks.
>
> 1. Ask and Guess - 5 minutes
>
> This task will give you a chance to show how good you are at asking questions. In the spaces on this page, write down all of the questions you can think of about the things you see in the picture. Ask the questions you would need to know to understand what is happening. Do not ask questions which can be answered just by looking at the picture. Ask as many questions as you can think of. If you need more space, use the other side of this sheet. You will have five minutes to work.

2. <u>Guess Causes</u> - 5 minutes

This task will give you a chance to demonstrate your ability to guess causes or formulate hypotheses concerning causations. In the spaces on this page, list as many possible things as you can which might have caused the action shown in the picture. You may use things which happened immediately before the action in the picture, or something that happened a long time ago. Make as many guesses as you can. Do not be afraid to guess as no one can really know from the information which you have. If you need more space, use the back of this sheet.

After five minutes, call time and go on to the third task. Guess Consequences.

3. <u>Guess Consequences</u> - 5 minutes

This task will give you a chance to show your skill in guessing the consequences of this action. In the spaces on this page, list as many possibilities as you can of what might happen as a result of what is happening in the picture. You may use things that might happen right afterwards or things that might happen as a result long afterwards in the future. Make as many guesses as you can. Do not hesitate to guess. No one can know for sure with the information which you have been given.

After five minutes call time and collect booklets.

The Ask-and-Guess Test was specifically timed for a total of 15 minutes, allowing 5 minutes for each part. Instructions for administering and the time limits imposed for each part of the Test were derived from Yamamoto (1962). Test-retest reliability correlations for these picture tests reported by Yamamoto (1962, p. 70) range from .56 to .83. Torrance (1963, p. 5) reports test-retest reliabilities from .79 to .89 with a total of .85 for all three parts of the test. Wodtke (1963, p. 56) reports test-retest reliabilities for the <u>Ask-and-Guess Test</u> ranging from .49 to .74 with an over-all of .66 on subjects at the fifth grade level.

All of the <u>A</u>, <u>B</u>, <u>C</u> tasks were timed with an imposed limit of 5 minutes for each. The instructions each teacher used in administering these three

were as follows:

> These are exercises to see how well you can use your imagination. Think of as many different ways as you can to complete the exercises in the time you have.
>
> Lost - Exercise A - 5 minutes
>
> You are to list what you would do or how you would use the things mentioned in the situation in order to feed and protect yourself.
>
> Rope Problem - Exercise B - 5 minutes
>
> Think of as many different ways in which you could use the given things in order to tie the ropes together.
>
> Rocket Ship Situation - Exercise C - 5 minutes
>
> List as many interesting things as you can think of that would happen (consequences) if this rocked ship was used by many people.

Samples of each of these three tasks are included in the Appendix A.

Scoring

All six criterion tasks for the experimental sample of 96 Ss were scored in the same procedure as the counterbalanced training word lists. Directions for scoring were prepared, and all Ss papers were coded, shuffled, and distributed among the same eight scorers used for the training tasks. The Ask-and-Guess Tests were scored for both originality and fluency. The same scoring criteria of selecting ten percent of the responses as original according to their infrequency within the sample population were adopted. The fluency score was the number of responses submitted by each S. A fluency score on these tasks was possible since each S was asked to think of as many responses as possible in the imposed time limit of 5 minutes for each part of the Ask-and-Guess test and for each of the A, B, C tasks. It should be mentioned, that all eight scorers

first viewed the Ask-and-Guess picture pertaining to the set of papers being scored. Sets of papers for each picture were scored separately.

Procedures and Design

Training

The experimental training program was conducted in sixth grade classrooms located in two schools. The writer served as the experimenter who presented the training task to all six classrooms over fourteen consecutive days of training. The three classes which served as the non-reinforced experimental groups which received practice only were trained first each morning between the hours of 9:15 and 10:00 in a consistent room order of 22, 23, and 24. The other three classes which comprised the reinforced experimental trained group were trained between the hours of 11:00 and 11:55, immediately following their morning recess period in a consistent room order of 10, 11, and 12.

Design

The experimental procedure of the study consisted of a counterbalanced design of both criterion tasks and training tasks across all six classrooms of the two experimental groups. A similar counterbalanced design was adopted for the criterion tasks given to the control group. Figure 8.01 is a representation of the counterbalanced design for the six criterion tasks for the three groups.

Fourteen days of training for creative performance were conducted by an experimenter across both experimental trained groups. Training task word lists were counterbalanced on the 7th, 11th, and 14th days of training.

Legend:

Task A - Lost Problem
Task B - Maier Rope Problem
Task C - Rocket Ship Situation
Task P_1 - Torrance Ask-and-Guess Picture Task (Old Mother Hubbard Picture)
Task P_2 - Torrance Ask-and-Guess Picture Task (Small Business Picture)
Task P_3 - Torrance Ask-and-Guess Picture Task (Cat in the Well Picture)

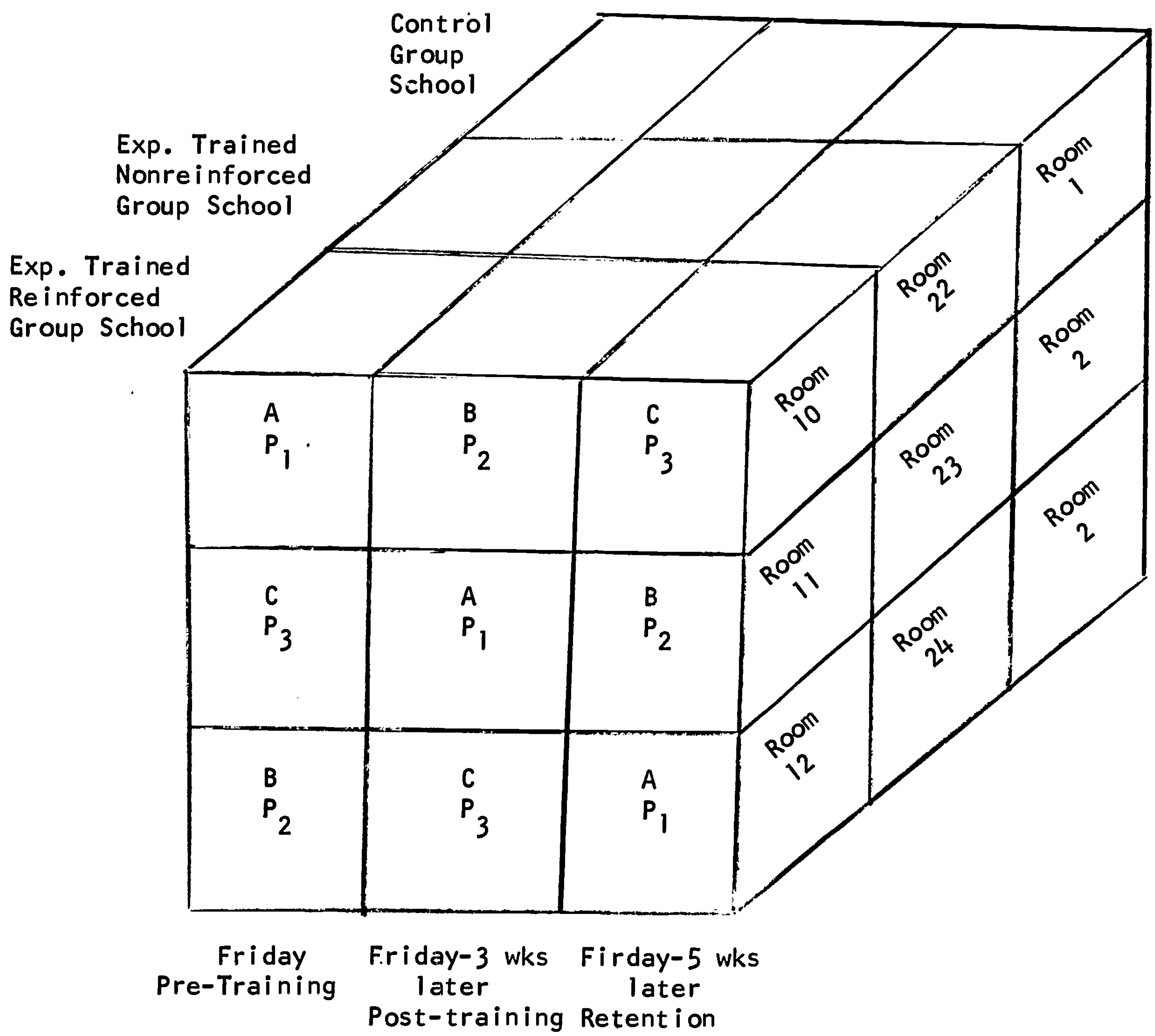


Figure 8.01. Representation of counterbalanced experimental design for criterion tasks.

Eleven other training word lists were repeated across all six classes for each of the other training days. A representation of the counter-balanced experimental design for the training task is shown in Figure 8.02.

Several weeks following the completion of training the experimenter again visited all six classrooms of the experimental trained groups. A form was distributed to each of the six classroom teachers asking them to choose their five most original and imaginative pupils. Their choice of the five were to be ranked from one to five as the most original child, next most original, and so on. Written instructions on the form were as follows:

> May we obtain your identification of the five most original and imaginative pupils in your class. These are not the grade-getters nor the most logical or critical thinkers, but those most creative and different in their thinking. As you evaluate your class, rank your most five creative thinkers as:

In this manner data on teachers choice of creative pupils was obtained so that a comparison could be made between those pupils recognized and selected by the teachers and those whose performance scores were the highest on the tasks selected to measure creative performance for the study.

RESULTS

If originality is a characteristic which is relatively stable from day to day in the amount elicited by a given task, then scoring procedures for the task could be tested for reliability. The reinforcement scoring of originality for each S on the even days of training (2nd, 4th,........ 14th) was compared with the originality scores on the odd days (1st, 3rd ....13th) of training. Comparison of the odd and even days scores

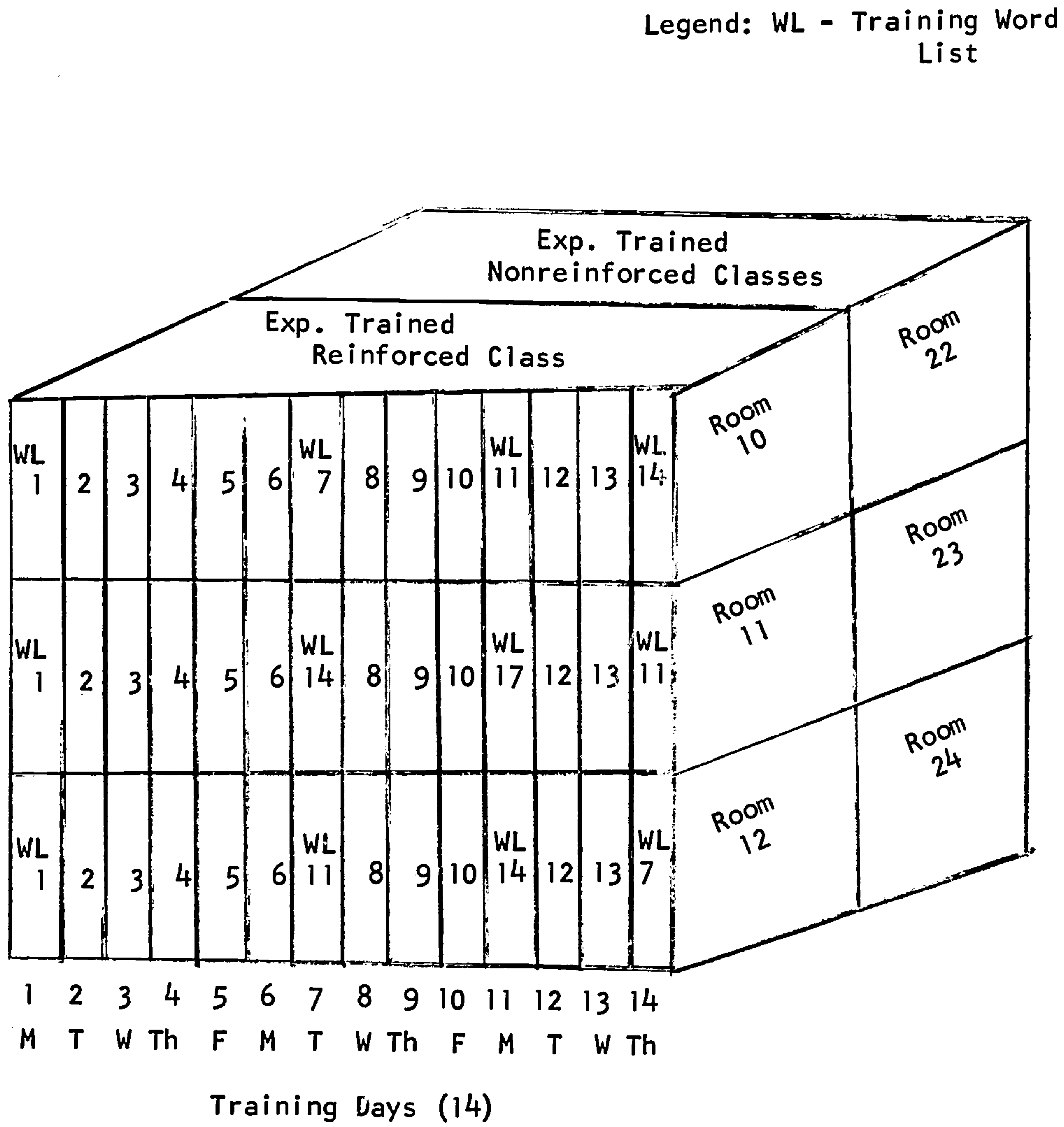


Figure 8.02. Representation of counterbalanced experimental design for training task.

of each S in Room 10 yielded a correlation (reliability coefficient) of .673 (N=16). The reliabilities of odd and even day originality scores of each S in Rooms 11 and 12 were .743 (N=17) and .608 (N=15). The overall reliability coefficient of the reinforcement scoring for the reinforced classes (N=48) of the experimental sample was .615. These reliability coefficients, considering the difficulty of scoring a variable such as originality, do show a considerable degree of reliability. This would in turn indicate that the trait designated "originality" as measured in this study appears as a relatively stable trait in a given S.

Although it was shown that the scoring of the reinforcement group was relatively reliable, it remained to be shown that the magnitude of those reliability coefficients was not just a function of the particular scorer, but that they were a reflection of the fact that such a measuring device could be scored reliably. The responses on the word list papers of the first day of training for both experimental groups were scored for originality by eight scorers. The papers were randomly assigned to eight different sets and each set scored by two scorers. The correspondence of the scoring of these two scorers for individual items was ascertained by the use of contingency coefficients. The contingency coefficients obtained were .31, .19, .26, .18, .11, .36 and .15 for the eight pairs of scorers, and .22 over-all. All were significantly different from zero ($p<.001$ for all except the contingency coefficient of .11 which was significant at the .05 level). Although the contingency coefficients are low numerically, it must be remembered that the highest possible contingency coefficient is .707 for two categories and this accounts in part for their small numerical size (Guilford, 1956, p. 316). Further, these coefficients

represent the reliability of the scoring for each item. The scores used in the subsequent analysis, however, were based upon two scorers.

Ten items per subject for each of the three counterbalanced training days were scored independently by each scorer. The number of responses selected as original by each scorer were summed to yield a total originality score for each S, and this score was used in the analyses of variance which follow. The scores obtained on each of the three days were summed to provide a total originality score that was used for the correlation analysis. A correlation of 0.48 was obtained by correlating the scores given to each case on the group of 30 items by one set of scorers and the scores given to each case on the same group of items by another group of scorers. The reliability of the two scores combined may be expected to be higher than this value and is estimated to be 0.69. The reliability of the originality score derived from the training task is also indicated by the fact that the score derived from the three counterbalanced days correlated 0.48 (N=96) with the originality score derived later from the Torrance Ask-and Guess test.

The Analysis of the Data

Analyses of variance were computed to determine the main effects of training, reinforcement, and transfer to the criterion tasks. A fixed model, 3 x 6 factorial design with repeated measures and unequal number of subjects per group was used in the analysis of variance of originality scores derived from training tasks. This same model was also used in the analyses of originality and fluency scores derived from the Ask-and-Guess and ABC criterion tasks for the experimental sample. A 3 x 3 factorial design with repeated measures and unequal number of subjects

per group was used in the analysis of fluency scores derived from criterion tasks for the control group. After each analysis of variance was completed the significant effects were studied using an individual degree of freedom test to determine which factors accounted for most of the variance. Each of these analyses with supporting data will be presented and discussed separately as follows:

Analysis of Variance for Counterbalanced Training Tasks of Both Experimental Groups

An analysis was made of the scores on originality across counterbalanced training task word lists by days and by classes. It was hypothesized that the simple training procedure employed in the experiment along with daily reinforcement would tend to facilitate original responses on the training tasks. An analysis of variance was carried out to test this hypothesis.

Table 8.02 presents the means and standard deviations of the data used for this analysis. The summary table for the analysis of variance of the originality scores on counterbalanced training days for both experimental groups is found in Table 8.03. As can be seen from the table, there were no significant effects. Thus, there is no increase in originality across training days. The other expectation that reinforcing original responses would result in more originality as evidenced by this training task was also not supported. There were no significant differences between classes within schools, and neither was there a difference between the reinforced and the nonreinforced classes.

TABLE 8.02

MEANS AND STANDARD DEVIATIONS FOR REINFORCED AND NONREINFORCED EXPERIMENTALLY TRAINED Ss ON ORIGINALITY FOR THE THREE COUNTERBALANCED TRAINING DAYS

		Training Days						
		7		11		14		
	N	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}\bar{x}$
Reinforced Classes								
10	16	2.81	2.37	3.44	2.48	2.63	1.96	2.96
11	17	3.35	2.18	1.71	1.65	2.47	1.94	2.51
12	15	2.00	1.65	1.60	1.50	3.20	2.34	2.27
Total	48	2.75		2.25		2.75		2.58
Nonreinforced Classes								
22	17	2.06	1.39	2.24	1.39	2.06	2.41	2.12
23	13	2.54	2.14	2.00	2.00	3.15	2.51	2.57
24	18	2.61	2.68	3.06	3.21	2.83	2.18	2.83
Total	48	2.40		2.50		2.64		2.51

TABLE 8.03

ANALYSIS OF VARIANCE FOR COUNTERBALANCED TRAINING TASKS

FOR THE TWO EXPERIMENTALLY TRAINED GROUPS

SCORED FOR ORIGINALITY

Source of Variation	Sum of Squares	df	Mean Squares	F	P
Between Subjects		95			
Classes	24.57	5	4.91	--	-
Schools (Reinf. vs. Nonreinforced	.12	1	.12	--	-
Class/Schools	24.45	4	6.11	--	-
Subjects/Classes	543.85	90	6.04	--	-
Within Subjects		192			
Days	6.97	2	3.49	--	-
Class x Days	54.52	10	5.45	1.35	-
Residual-subjects					
Within groups	727.52	180	4.04	--	-
Total	1357.43	287			

Analysis of Variance for the Counterbalanced Ask-and-Guess Criterion Task of Both Experimentally Trained Groups Scored for Originality

The reader will recall that the Torrance Ask-and-Guess Test was scored for both originality and fluency. Both of these abilities as factors of creative performance were analyzed separately since they yielded two distinct scores. Two analyses of variance were completed for testing differences between classes and days in terms of the Ask-and-Guess tasks which were presented before training, after training and two weeks following training. The hypothesis was that Ss would learn to respond with a greater number of original ideas as a result of training and such learning would transfer from the training task to the criterion tasks. An analysis of variance was completed separately for originality and fluency scores to test this hypothesis. In Table 8.04 are the means and standard deviations for the originality scores obtained from the Ask-and-Guess Test. The summary table for the analysis of variance for these data is shown in Table 8.05. As can be seen from the table, class differences within schools were significant. This effect can be seen from Table 8.06 to be largely a result of the fact that rooms 10 and 23 had substantially higher scores than the other rooms. The reinforced group was not different from the nonreinforced group classes.

The effect of days was not significant. This shows once again that training and/or reinforcement does not increase the amount of originality displayed by the Ss. It should be repeated that the across days comparison here involved a pre-training task, a post-training task, and a retention task given two weeks after the post-training task in which it becomes evident that the amount of originality displayed by the groups as a whole

TABLE 8.04

MEANS AND STANDARD DEVIATIONS FOR REINFORCED AND NONREINFORCED EXPERIMENTALLY TRAINED Ss ON THE ASK-AND-GUESS CRITERION TASK ADMINISTERED BEFORE, AFTER AND TWO WEEKS FOLLOWING TRAINING SCORED FOR ORIGINALITY

		Time of Administration						
		Pre Training		Post Training		Two Weeks Following Training		
	N	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\overline{Xx}$
Reinforced Classes								
10	16	6.88	2.82	9.94	3.30	9.50	7.67	8.77
11	17	6.24	3.53	6.59	1.79	7.06	3.03	6.63
12	15	6.40	3.70	6.00	2.62	5.93	3.99	6.11
Total	48	6.50		7.52		7.51		
Nonreinforced Classes								
22	17	6.06	3.77	6.59	4.86	6.65	5.30	6.43
23	13	8.15	3.08	9.15	3.48	9.38	5.99	8.89
24	18	7.44	4.12	6.72	3.94	7.28	2.63	7.30
Total	48	6.94		7.33		7.63		

TABLE 8.05

ANALYSIS OF VARIANCE FOR COUNTERBALANCED DAYS AND CLASSES OF THE TWO EXPERIMENTALLY TRAINED GROUPS ON THE ASK-AND-GUESS CRITERION TASK SCORED FOR ORIGINALITY

Source of Variation	Sum of Squares	df	Mean Squares	F	P
Between Subjects		95			
Classes	346.30	5	69.26	2.91	.05
Schools (reinf. vs. Nonreinforced)	.35	1	.35	----	---
Class/Schools	345.95	4	86.49	3.63	.05
Subjects/Classes	2142.00	90	23.80	----	---
Within Subjects		192			
Days	30.54	2	15.27	1.16	---
Class/Days	81.84	10	8.18	----	---
Residual -- subjects Within Groups	2360.00	180	13.11	----	---
Total	4960.68	287			

TABLE 8.06

MEANS AND STANDARD DEVIATIONS FOR REINFORCED AND NONREINFORCED EXPERIMENTALLY TRAINED Ss ON THE ASK-AND-GUESS CRITERION TASK ADMINISTERED BEFORE, AFTER AND TWO WEEKS FOLLOWING TRAINING SCORED FOR FLUENCY

		Time of Administration						
		Pre Training		Post Training		Two Weeks Following Training		
	N	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}\bar{X}$
Reinforced Classes								
10	16	30.81	12.65	46.06	15.39	33.56	14.71	36.81
11	17	24.94	9.27	30.59	12.91	37.12	19.63	30.88
12	15	28.07	11.57	34.67	12.55	40.33	20.13	34.36
Total	48	27.87		37.02		36.94		
Nonreinforced Classes								
22	17	20.71	10.05	24.12	11.32	22.35	15.04	22.39
23	13	27.08	11.87	37.23	16.98	35.92	21.17	33.41
24	18	34.06	16.18	30.94	11.12	44.83	12.43	36.61
Total	48	27.44		30.23		34.46	.	30.71

is remarkably stable. In fact, the indication is that originality may be a very stable trait which is little affected by direct training or reinforcement. This does not exclude the possibility that some forms of indirect training may not be effective.

<u>Analysis of Variance for the Counterbalanced Ask-and-Guess Criterion Task of Both Experimental Groups Scored for Fluency</u>

Table 8.06 shows the means and standard deviations for this analysis. The summary Table 8.07 of the analysis of fluency for the <u>Ask-and-Guess</u> criterion test indicates the following: When the fluency scores of the two experimental groups (reinforced and nonreinforced classes) on the <u>Ask-and-Guess Task</u> were compared, the main effects of classes ($p<.05$) and days ($p<.001$) were significant as well as the interaction ($p<.001$). When the between classes effect was broken apart, no difference was found between the reinforced group and the non-reinforced group. The variance related to this effect was quite widely distributed across the various classes and no generalizations could be made.

In view of the findings derived from the other two analyses of variance that training did not affect the amount of originality displayed by the Ss; the above finding that fluency (total number of responses) did increase across days indicates a practice effect on the criterion tasks independent of training since fluency was not required on the training task. This suggests that originality may be a very stable trait which cannot be altered by this kind of training; and which does not display a practice effect. Even though Ss can benefit from practice and can learn to respond more fluently, the data indicate that they apparently cannot

TABLE 8.07

ANALYSIS OF VARIANCE FOR COUNTERBALANCED DAYS AND CLASSES OF THE TWO EXPERIMENTALLY TRAINED GROUPS ON THE ASK-AND-GUESS CRITERION TASK SCORED FOR FLUENCY

Source of Variation	Sum of Squares	df	Mean Squares	F	p
Between Subjects		95			
Classes	6819.8	5	1363.9	3.05	.05
Schools (Reinf. vs. Nonreinforced	250.3	1	250.3	--	--
Classes/Schools	6569.5	4	1642.4	3.68	.05
Subjects/Classes	40211.7	90	336.8	--	--
Within Subjects		192			
Days	3417.8	2	1708.9	19.80	.001
Tasks	1541.1	2	770.6	8.65	.001
Interaction	1735.5	8	217.1	2.44	.025
Residual Within Subjects	16041.0	180	89.1	--	--
Total	69767.9	287			

be trained to be more original. The significant interaction (P=.025 level of confidence) may suggest that something other than supposedly controlled experimental conditions was causing the interaction. There was a considerable amount of variation with respect to which of the three days produced the highest fluency for each room. For example, with Room 10 the post test produced the most responses, with Room 11 Ss the later test produced the most responses, and so on. This suggests that incorrect timing could be the cause since these were regular classroom exercises given by the regular teacher and integrated within the program of the school day.

<u>Analysis of Variance for the Counterbalanced ABC Criterion Tasks of Both Reinforced and Nonreinforced Groups Scored for Fluency</u>

With the preceding evidence which revealed no significant differences in originality scores as shown by the analyses of variance for the counterbalanced training tasks and the Ask-and-Guess criterion task across days and classes it was hypothesized that it would be unlikely for any differences to occur on originality across the other teacher controlled ABC criterion tasks. In order to relieve the scoring burden, these teacher controlled criterion tasks were only scored for fluency by a single scorer. An analysis of variance was completed in order to again test the effects of practice and transfer from an experimenter training task to a teacher controlled classroom task calling for a fluent number of responses. The means and standard deviations for these data are shown in Table 8.08. The analysis of variance for the data on the ABC criterion tasks for the experimental groups scored for fluency is presented in Table 8.09.

TABLE 8.08

MEANS AND STANDARD DEVIATIONS FOR REINFORCED AND NONREINFORCED EXPERIMENTALLY TRAINED Ss ON THE "ABC" CRITERION TASKS ADMINISTERED BEFORE, AFTER AND TWO WEEKS FOLLOWING TRAINING SCORED FOR FLUENCY

		Time of Administration						
		Pre Training		Post Training		Two Weeks Following Training		
	N	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\overline{Xx}$
Reinforced Classes								
10	16	4.50	1.51	8.25	5.89	11.56	7.11	8.10
11	17	10.29	4.12	4.59	1.66	7.88	3.85	7.59
12	15	9.87	3.50	8.73	2.60	3.87	0.52	7.49
Total	48	8.23		7.10		7.85		7.73
Nonreinforced Classes								
22	17	8.71	4.37	7.41	4.54	4.29	1.65	6.80
23	13	3.92	0.64	11.08	3.71	8.69	2.92	7.90
24	18	6.00	1.97	3.83	0.51	10.28	5.14	6.70
Total	48	6.40		7.06		7.73		7.06

TABLE 8.09

ANALYSIS OF VARIANCE FOR COUNTERBALANCED DAYS AND CLASSES OF THE TWO EXPERIMENTALLY TRAINED GROUPS ON THE "ABC" CRITERION TASKS SCORED FOR FLUENCY

Source of Variation	Sum of Squares	df	Mean Squares	F	p
Between Subjects		95			
Classes	77.13	5	15.43	--	--
Schools (Reinf. vs. Non reinforced	10.77	1	10.77	--	--
Classes/Schools	66.36	4	16.59	--	--
Subjects/Classes	2027.70	90	22.53	--	--
Within Subjects		192			
Days	16.84	2	8.42	--	--
Tasks	1524.26	2	762.13	8.66	.001
Interaction-Tasks x Days	425.45	8	53.18	6.04	.001
Residual--subjects Within Groups	1584.00	180	8.8	--	--
Total	5655.38	287			

The effects of classes and days were not significant. There were no significant differences in the mean fluency scores on these tasks between classes or across the three testing occasions. The significant interaction ($p<.001$) may be the result of one task (B) being significantly more difficult than the other two. Table 8.10 presents means for all six classes arranged by tasks. Means by tasks indicate that task B was essentially twice as difficult as the other two. Since this task was counterbalanced across days with respect to classes, the fact that it was more difficult could produce the interaction effect. The Latin square used in the counterbalancing of tasks and days hardly permits the discovery of an interaction which has much meaning.

Analysis of Variance for the Counterbalanced Ask-and-Guess Criterion Task of the Control Group Scored for Fluency.

A control group was used to measure any training effect that might be attributable to the taking of the tasks themselves. The Ask-and-Guess criterion task was conducted over the same time interval for the control group as was the testing for experimental groups; that is, over one, three and five weeks time. Because fluency scores for the experimental groups were significantly different across the three days on the Ask-and-Guess task, the control group was scored for fluency on the same task to see if differences would be due to training. The question that remained to be answered was whether training affected fluency even though it didn't include a fluency task. An analysis of variance was completed to test if there were any significant differences between counterbalanced tasks on the three testing days. Table 8.11 gives the means and standard

TABLE 8.10

MEANS FOR THE COUNTERBALANCED "ABC" CRITERION TASKS OF THE TWO EXPERIMENTALLY TRAINED GROUPS SCORED FOR FLUENCY ARRANGED BY TASKS

Classes	Task A	Task B	Task C
10	8.25	4.50	11.56
11	7.88	4.59	10.29
12	9.87	3.87	8.73
22	8.71	4.29	7.41
23	11.08	3.29	8.69
24	10.28	3.83	6.00
Totals	56.07	25.00	52.68

TABLE 8.11

MEANS AND STANDARD DEVIATIONS FOR CONTROL GROUP Ss ON THE ASK-AND-GUESS CRITERION TASK ADMINISTERED OVER THE SAME TIME INTERVAL AS THE EXPERIMENTALLY TRAINED GROUPS SCORED FOR FLUENCY

		Time of Administration						
		First Week		Third Week		Fifth Week		
Classes	N	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}x$
1	11	26.55	8.72	22.82	8.61	34.82	17.73	28.06
2	15	19.73	7.31	23.13	8.04	22.87	11.51	21.91
3.	12	33.83	11.13	30.67	7.66	36.42	12.32	33.64
Totals	38	26.16		25.42		30.61		27.40

deviations for the scores. As can be seen from Table 8.12, the effects of days and classes are significant ($p<.001$ and $p<.05$ respectively). Because of the IQ differences across classes, the effect of classes was broken apart and it was found that the differences between the lowest mean IQ class and the highest mean IQ class was not significant ($F=2.35$, 4.13 needed for the $P=.05$ level of confidence).

The effect over time (days) is obviously accounted for in terms of the difference between the first two days and the last. The gain suggests that a similar gain found in the case of the trained group is not a result of training but is probably due to some facilitation resulting from the taking of the tests.

Studies Relating Measures of Originality and Fluency to Measures of Pupil Performance

Correlations of Intelligence with Originality and Fluency

Table 8.13 indicates correlations of The *California Mental Maturity Scale* with measures of originality and fluency scores on all tasks administered the two experimentally controlled groups. The correlations of this measure of intelligence with originality as measured by the training task, are shown in Table 8.13. Although with only one exception the correlations by rooms were not significant ($r=.51$, $p<.05$ for Room 24), the correlations by room were all positive and did indicate a trend for the two factors to be positively correlated. When the scores were not separated by room but were considered altogether, the correlation was significant ($r=.27$, $N=96$, $p<.05$). One could conclude that there was a significant correlation between intelligence (*California Mental Maturity*

TABLE 8.12

ANALYSIS OF VARIANCE FOR COUNTERBALANCED DAYS AND CLASSES OF THE CONTROL GROUP ON THE ASK-AND-GUESS CRITERION TASK SCORED FOR FLUENCY

Source of Variation	Sum of Squares	df	Mean Squares	F	p
Between Subjects					
Classes	2563.08	2	1281.54	4.90	.05
Subjects within groups	9193.00	35	262.70	--	--
Within Subjects					
Days	709.28	2	354.64	12.64	.001
Interaction	41.60	4	10.40	--	--
Residual--subjects Within Groups	1964.00	70	28.06	--	--
Total	5470.96	113			

TABLE 8.13

CORRELATIONS OF INTELLIGENCE WITH ORIGINALITY AND FLUENCY SCORES ON THE TRAINING AND CRITERION TASKS FOR THE TWO EXPERIMENTALLY TRAINED GROUPS

Room	N	Originality: Training Task	Originality: Ask-and-Guess Criterion Task	Fluency: Ask-and-Guess Criterion Task	Fluency: "ABC" Criterion Task
Reinforced					
10	16	.30	.29	.18	.38
11	17	.45	.16	.46	.04
12	15	.04	.26	.38	.18
Nonreinforced					
22	17	.15	.02	.24	.00
23	13	.30	.23	.09	.24
24	18	.51*	.27	.00	.03
Overall	96	.27*	.28*	.23*	.13

r's converted to Fisher's Z functions, averaged and reconverted to rv.

*$p<.05$

Scale) and originality as measured by this task for the two groups together.

The correlations of intelligence with originality on the Ask-and-Guess criterion task by rooms were positive (with one exception for Room 22) and while individually nonsignificant, were collectively significant. In order to obtain an over-all correlation for all classes it was necessary to combine them in such a way that class differences would not affect the relationship since it was suspected that some teachers may have allowed excessive time in administoring the Ask-and-Guess Test. This was done by converting them to Fisher's Z function, averaging the values, and reconverting them to an over-all correlation. This was done for all correlations on individual classes. When this correlation was made the over-all correlation of originality measured by the Ask-and-Guess criterion task with intelligence was .20 which was significant ($p<.05$).

The fluency scores on the Ask-and Guess criterion task correlated with intelligence. These correlations varied from moderate positive to zero. Once again, in order to eliminate any class differences which may have resulted from differences in timing the individual room correlations were converted to Fisher's Z function, averaged, and reconverted to an over-all correlation. An over-all significant correlation of .23 was obtained ($p<.05$).

Fluency scores derived from the ABC criterion task were also correlated with intelligence test scores. For all rooms, these were all positive and low with the exception of one room (22) which was zero. The correlations for all rooms were again converted to Fisher's Z function, averaged and reconverted to an over-all correlation of .14 which was not significant.

An over-all correlation of intelligence with fluency as measured by the Ask-and-Guess criterion task for the control group was .51. This would lead to the conclusion that this measure of fluency was correlated with intelligence ($p<.01$).

Correlations of Training Task Originality Scores Correlated with Teacher Grades, Citizenship, Study Habits and Grade Placement on a Standardized Achievement Test

Correlations between teacher academic grades, citizenship grades, study habits grades, standardized achievement test scores and originality and fluency scores on the counterbalanced Ask-and-Guess Test for each experimental group are presented in Table 8.14.

Several points may be made concerning the correlations obtained from these data. It is apparent from the significant correlations of originality measures derived from scores on both the training tasks and the Ask-and-Guess task with the measures of pupil performance that these teachers do not discriminate against the original, imaginative child. In fact, the data indicate that there is a positive relationship between grades, citizenship and study habit marks given by teachers on the one hand, and the student's originality and fluency performance on the measures used in the experiment. It is interesting to note the size of correlations (.31 and .37) obtained between mathematics, science and social studies grades with originality and fluency measures on the Ask-and-Guess task. In addition, the correlation of .375 was found between these subject matter grades and the training task originality scores. The fact that the correlation of originality with language arts, art and music grades is

TABLE 8.14

CORRELATIONS OF ORIGINALITY AND FLUENCY ON THE ASK-AND-GUESS CRITERION TASK WITH TEACHER GRADES, STUDY HABITS, CITIZENSHIP AND STANDARDIZED ACHIEVEMENT TEST SCORES

		Originality									
Room	N	(Math-Sci-SS)	Grades (Lang.-Art-Music)	Citizen-ship	Study Habits	Ach. Test	Grades (Math - Sci-SS)	Grades (Lang.-Art Music)	Citizen-ship	Study Habits	Ach. Test
10	15	.30	.41	.21	.15	.48	.15	.09	.00	-.08	.31
11	17	.28	.50	-.17	.24	.21	.60	.18	.00	.48	.22
12	14	.24	.55	.43	.43	.50	.26	.25	.21	.32	.18
22	16	.45	.44	.51	.33	.09	.57	.72	.67	.48	.08
23	11	.52	.56	.39	.60	.52	.61	.58	.27	.44	.13
24	17	.01	-.02	-.29	-.26	.21	.16	-.19	.00	-.35	-.02
*Total		.31	.42	.20	.27	.42	.37	.31	.21	.23	.15

*r's converted to Fisher's Z Function, averaged and reconverted to r.

even higher (.42) as measured by the <u>Ask-and-Guess</u> task indicates that these teachers recognize and rewarded creative students. The higher correlation of originality with language, art and music might be expected since these subjects, especially language arts, perhaps deal more with the same kinds of creative abilities as do the written tasks used in the experiment for measuring originality. One of the most significant findings is the size of the correlations of both originality and fluency as measured by the Ask-and-Guess tests with citizenship and study habit marks. These correlations certainly indicate that these teachers do not discriminate against students who measure high on originality and fluency. Another most important speculation is made when observing the significant correlations of originality (.42) as measured by the <u>Ask-and-Guess</u> criterion task with grade placement as determined by a standardized achievement test score. This correlation, along with those of originality with subject matter grades, could indicate that knowledge may be a key factor in the functioning of creative abilities. Original responses generally require a breadth of information and experience. At least it shows that the student who has sufficient knowledge to place at a higher grade level than where he is now also performs better on creativity measures involving original thinking. This observation is even more interesting when observing the correlations of achievement test grade placement with training and criterion measures for originality (.42 and .31) as compared with fluency (.15). Apparently the actual knowledge of subject matter content has a higher relationship with original thinking than the ability to merely provide a stream of relevant responses regardless of their originality. The higher originality correlations across all of the measures of pupil

performance as compared with fluency correlations may indicate additional evidence to support this position.

The last observation which should be pointed out when interpreting the resulting correlations in Table 8.14 is that of individual room comparisons. With the exception of a few scattered deviant correlations, most of the teachers were fairly consistent except one, Room 24. It should be noted that this teacher apparently places a different emphasis on citizenship and study habit marks than the other five teachers.

Teacher's Selection of Original and Imaginative Students Related to the Ask-and-Guess Originality Scores

It should be recalled from the chapter discussing the experimental method and procedure that all six teachers of the experimental classes were asked to select and rank their five most original and imaginative students. The criterion for an original and imaginative student was explicitly given in written form to each teacher as "not necessarily the grade-getter nor the most logical, but the student most creative and different in his or her thinking."

Originality score derived from the Ask-and-Guess criterion task of those students so selected by each classroom teacher was then compared by classes with originality scores derived from the same measure by the remaining students in the experimental sample not so selected by the teachers. Table 8.15 presents these data for each of the six classrooms of the experimental sample. Only one classroom teacher's selection (room 10) of his most original and imaginative students approached

TABLE 8.15

COMPARISON OF MEAN ORIGINALITY SCORES BY CLASS DERIVED FROM THE ASK-AND-GUESS CRITERION TEST FOR THOSE STUDENTS SELECTED BY EACH CLASSROOM TEACHER AS MOST ORIGINAL AND THE REMAINING STUDENTS NOT SELECTED

Room	$Mean_1$ Selected by Teacher	$Mean_2$ Not Selected	Differences	t	df
10	35.00	23.42	11.58	2.05*	15
11	17.25	20.69	- 3.44	0.91	16
12	20.50	17.55	2.95	0.81	14
22	24.33	18.21	6.12	0.98	16
23	26.00	26.90	- 0.90	0.13	12
24	20.50	21.71	- 1.21	0.33	17

significance ($p<.05$) when t tests were computed. All of the other five classroom teachers were unable to select their five most original students at any significant level of confidence as determined by the originality measure derived from the Ask-and-Guess test. These data may be interpreted to indicate that whatever the classroom teachers' criterion for an original student may be, it is apparently not the same as that which is being measured by the Ask-and-Guess Test. In essence, at least five of the six teachers were not selecting the same students on the basis of originality that the test measure did. This data may indicate that there is little agreement between teacher's judgment of originality and this particular test's measure of supposedly the same trait. Just why this disagreement occurs cannot be determined from these findings.

Discussion

This study was undertaken with considerable scepticism concerning the trainability of original responses. While the Maltzman studies had generally shown that practice in the generation of original responses showed no transfer effect to performance on the Guilford Unusual Uses Test, one could argue that little transfer might have been produced owing to the very limited amount of training given. The hope of the present study was that with more extended practice a training effect might appear. In this respect, the data derived from the present study is unequivocal for there was not the slightest sign that training had any effect on the development of a trait of originality.

Perhaps the results of the present study are to be expected. First, it is very clear from an examination of the responses produced to the training task that the production of original responses is highly related to the amount of information possessed by the individual. A child who is asked to find an original use for a pencil and who suggests that the carbon might be used in building an atomic furnace is a highly knowledgeable youngster. The limit on original responses would appear to be the amount of stored knowledge of the person who is attempting to make the response. Practice in the making of original responses is hardly likely to increase the amount of knowledge stored. There is still the possibility that practice may somehow permit the individual to learn how to bring one stored item of information into a relationship with another stored item of information. Some comment on this possibility is in order.

Most modern psychological theories take the position that a particular item of information is not stored in a particular location in the brain, but rather that there is a redundancy of storage. Considerable physiological evidence supports such a position. The result of this is that there is a built-in-relationship between any item of information and a vast range of other items of information. Associationist theories generally assume that these relationships can be represented by a hierarchy of associations of different strength. Such a hierarchy of relationships involving even vastly different items of information is possible because of the high redundancy of the storage system. Associationistic theory also generally takes the position that the associations between items of information is a result either of relationships established through circumstances of learning or that the relationships may exist through the

accidents of storage which tend to bring one item of information into relationship to another. At no place in such a theory is the idea introduced that such associations can be developed by some kind of training which develops skill in locating one item of information in the brain and then finding another item with which it has not been associated and then, somehow, tying the two together. From the point of view of associationistic theory, as well as any other aspect of learning theory, this notion is plain nonsense. The associations available are presumably built into the brain.

The associationistic view does not imply that all associations between concepts or elements of information can be manifested at any time. Some of the weaker associations can function only on rare occasions or under exceptional circumstances. Hebb (1949) has pointed out that accidental events, such as the simultaneous arrival of two nerve impulses at the same point at the same time, may trigger neural events which would not otherwise occur. Such accidental triggering would appear to play a role in the production of remote associations. The only externally controllable condition which one can identify which might facilitate the production of such remote associations would be the inhibition of common associations. Thus the notion that practice in the giving of remote associations will facilitate such an activity seems to be psychologically very unlikely. The educational theory that the basic elements of creative talent, other than the cognitive, are trainable has emotional appeal but little scientific status.

The correlations between the measures of originality and fluency and measures of educational performance are interesting. They give no

support to the reckless statements which have been made to the effect that teachers generally discriminate against the creative child. In only one class out of the six main experimental classes was there the slightest evidence that this was so. The child who scores high on a test of originality is likely to have higher than average grades and is likely to be better informed than the average child. In the class that is the exception in this study, the teacher involved was quite hostile to the study and was observed to have an unusually severe and authoritarian approach towards children. Our observations in schools lead us to believe that the teacher-child relationship in this particular class was an unusual one. The evidence from the study is consistent with the position that the original child, who inevitably brings so many ideas of interest to a class, is generally favored by the teacher. Teachers, like children, enjoy elements that bring freshness to the schoolroom atmosphere.

Of interest are the correlations between originality scores and achievement as measured by standardized tests. Such tests reflect to a considerable extent the amount of information stored by the student from the academic areas which the tests cover. The correlations give support to the position already voiced that a crucial element in the production of original responses is knowledge. The latter position seems to be as sound as any that has been taken in the area and leads to the implication that one of the best ways of training the ability to produce original responses is to provide the pupil with a rich background of knowledge. It may well be that the acquisition of knowledge may have a far more powerful effect on originality than all of the training procedures which have been devised and which have such superficial attractiveness.

References

Anderson, R. C., and Anderson, R. M. Transfer of originality training. J. educ. Psychol., 1963, 54, 300-304.

Christensen, P. R., Guilford, J. P., and Wilson, R. C. Relations of creative responses to working time and instructions. J. exp. Psychol., 1957, 53, 82-88.

Gerlach, V. S., Schutz, R. E., and Baker, R. L. Effect of variations in test directions on originality test response. J. educ. Psychol., 1964, 55, 79-83.

Knight, D. A. Fluency, flexibility, and originality training as related to creativity. A thesis submitted to the faculty of the University of Utah in partial fulfillment of the requirements for the degree of Master of Science, 1962.

Maltzman, I., Bogartz, William, and Breyer, L. A procedure for increasing word association originality and its transfer effects. J. exp. Psychol., 1958, 56, 392-398.

Maltzman, I., Simon, S., Raskin, D., and Licht, L. Experimental studies in the training of originality. Psychol. Monog., 1960, 74, Whole No. 493, p. 23.

Meadow, A., and Parnes, S. J. Evaluation of training in creative problem solving. J. appl. Psychol., 1959, 43, 189-194.

Meadow, A., Parnes, S. J., and Reese, H. Influence of brainstorming instructions and problem sequence on a creative problem solving test. J. appl. Psychol., 1959, 43, 413-411.

Parnes, S. J., Effects of extended effort in creative problem solving. *J. educ. Psychol.*, 1961, 52, 117-122.

Parnes, S. J., and Meadow, A. Effects of "brainstorming instructions on creative problem solving by trained and untrained subjects. *J. educ. Psychol.*, 1959, 50, 171-176.

Parnes, S. J., and Meadow, A. Evaluation of persistance of effects produced by a creative problem-solving course. *Psychol. Rep.*, 1960, 7, 357-361.

Torrance, P. E., and Gowan, J. C. *The reliability of the Minnesota Tests of Creative Thinking.* St. Paul, University of Minnesota, Bureau of Educational Research, 1963.

Wodtke, K. H. *A study of the reliability and validity of creativity tests at the elementary school level.* Doctoral Dissertation in the library of the University of Utah, 1963.

Yamamoto, Kaoru. *Revised scoring manual for Tests of Creative Thinking.* St. Paul, University of Minnesota, Bureau of Educational Research, 1962.

CHAPTER IX

A REVIEW OF SECTION II

The first seven chapters in Section II form a series of closely related studies with findings which can be considered collectively. The eighth chapter presents a different kind of problem - the application of reinforcement to a category of behavior which is not likely to be strengthened by such a procedure. While the reinforcement principle has powerful applications to many problems of teaching, the last of the series of studies presented here was undertaken with the idea that certain behaviors are not facilitated or increased in frequency through reinforcement but must be developed by other procedures.

The first series of studies all involve a rote learning task somewhat similar to that observed frequently in classrooms. The general format of the task remained the same throughout the various investigations. The acquisition task was that of learning the meaning of 60 German words during three periods which were always scheduled on Monday, Tuesday, and Wednesday of each week. The task was chosen because few children in the grades studied had been exposed to German and, hence, previous learning on the part of some children was not likely to be a factor contaminating some of the data. The general form of the task was that of presenting a German word together with two English words. The task of the subject was to determine which one of the English words had a meaning equivalent to that of the German word. The task could be presented under a number of different conditions and the studies involved a determination of the relationship of the learning conditions to the degree to which the pupils learned the task. The level of acquisition was always measured on Friday, but in one study a delayed retention test was also administered some weeks later.

The first study in the series, presented in Chapter II, investigated

the acquisition of the German vocabulary under conditions similar to those which commonly occur in classrooms. The situation involved eight pupils sitting in a row and an experimenter-teacher who presented the German words and the two English words on large cards. The teacher interacted with some of the pupils but not with others. When the pupil chose the correct alternative, he was told by the experimenter that he was right. When he chose the wrong alternative, the teacher said nothing and turned to another pupil with the next problem. This plan of reinforcement was selected because it had been observed previously that teachers typically say nothing when the response of the pupil is incorrect. In addition, some pupils learned the task from a teaching machine. Under one of the teaching machine conditions the pupil worked alone. In the other the pupil and the experimenter sat side by side and the experimenter said either "right" or a synonym for right when the pupil made a correct response.

The rate of acquisition showed a typical learning curve with the strength of the acquired response being related to the number of trials with the particular item. In this study, those who had direct interaction with the teacher in the familar recitation-type of learning situation were those who learned most effectively (on the items on which interaction occurred.) However, when some interaction was provided in a teaching machine type of learning situation a similar advantage was not observed. Indeed, when the feedback was provided both by the experimenter and by the machine a particularly inefficient learning situation was observed to exist.

Those who learned by observing the other pupils did not perform as well on the achievement test as those that learned through a situation in which there was interaction with the teacher in a group situation. Their

performance was similar to the performance of the interacting group on those items which the interacting group had to learn by observation. This suggests that the advantage gained by the interacting group is confined to those items on which there is interaction with the teacher. In the interacting group there was no general raising of the level of the performance.

The first study brings out clearly that, among the conditions studied, the pupil who responds orally and who interacts with the teacher learns best, and the pupil who plays a strictly passive role learns least.

The second study of the series reported in Chapter III explores the effect of the form of the teacher's feedback on the achievement of the pupils in a recitation rote-learning situation of the kind used throughout the studies. In this study the form of the feedback was varied and also the amount and redundency of the information transmitted. The general finding of this study is that the more complex the feedback and the less the use of the feedback is dependent upon the memory of the learner the more effective it is. The most effective feedback was that which repeated the essentials of the problem together with the right answer in the form "(German word) means (English word.)" This form of feedback involved redundency and also eliminated the need for the learner to remember or review the problem for himself. In this second study, the form of feedback which was most efficient for those pupils interacting with the teacher, was also most efficient for those learning by observation. The data suggest that whatever conditions make for effective learning of subjects, who are both orally responding and interacting, are also the conditions which make for the most effective learning of those who learn by observation. Another significant finding of this study is that the subjects who interacted

with the teacher showed less decline in learning over days than those who did not interact. This finding has come out in a number of places in our data and, in general, those who have learned alone without social interaction have shown substantially more decline over the three days than have the children working in situations where a greater amount of social interaction occured. This study, unlike the previous study produced some evidence to indicate that a pupil who responded on one item not only learned that item more effectively than did the pupil who did not respond, but also tended to learn more effectively the other items on which he did not respond.

The experiment on the effectiveness of different kinds of feedback also provided data pertaining to the relationship of the judged attention of the student to the amount of learning taking place. The judgments placed heavy emphasis on the direction of the gaze and the general appropriateness of the body orientation. These aspects of behavior were found to be relevant for judging the extent to which pupils were learning in those situations in which the orientation of the subject was necessary for the acquisition of information. For example, if the subject had to read the words on a card in order to learn the task, then the orientation of the subject towards the reading material is clearly a necessary condition for learning. However, under those conditions where all the information necessary for learning is transmitted through the auditory channel, then attention as it is judged in this study and as it is ordinarily judged is unimportant for acquisition of the material.

Chapters V and VI present studies in which pupils work with pupils and in which the responses of pupils are reinforced by other pupils rather than by teachers. This approach to teaching is an old one though there have been almost no data available concerning the effectiveness of learning

situations in which pupils teach pupils. In the study reported in Chapter V, elementary school pupils in grades 4, 5, and 6, learned under a number of conditions in which pupils assumed both the role of teacher and the role of pupil. In summarizing the results of this study and to prevent any unwarranted overgeneralization, the point must be made that the materials were carefully developed so that one pupil could easily fall into the role of teacher and the other could readily assume the role of pupil. Unless materials have been very carefully prepared and are such that they provide considerable structure for the behavior of the children involved, effective learning is unlikely to take place.

Among the various conditions of learning investigated in this research, the least effective was that in which the pupil served as teacher. However, the teacher-pupils did learn substantial amounts despite the fact that their role does not provide optimum conditions for learning. On total learning scores, the pupils who responded to each item and who were reinforced by the teacher pupil obtained the highest average scores. The second highest average scores were obtained by pupils who worked through the deck of cards alone. The latter pupils also obtained average scores comparable to the pupils in a previous study who learned the same task by means of a teaching machine. Switching the pupil and the teacher role in the middle of the task did not have the effect of vitalizing learning in the manner that had been anticipated. Those who switch roles obtained mean scores almost exactly between those who acted throughout the entire session in a teacher role and those who performed throughout in a pupil role.

A delayed test measuring retention was administered at varying

intervals after the time of acquisition. The most striking feature of the data on delay is the remarkably low loss over the 9 week period. This may be due to the fact that the task was unrelated to school learning and, hence, new material learned in school during the period would be unlikely to have a disruptive effect on the learning undertaken as a part of the experiment. The data fit well an interference theory of forgetting in that such a theory would predict that in this case little forgetting would occur.

An efficiency score was computed for each learning condition. This score was the time taken divided by the number of items answered correctly and thus could be considered as a measure of time per item learned. The group which had by far the highest efficiency scores were those who worked alone since they required only about half of the time utilized by other subjects in order to complete the tasks. However, the advantage of the isolated conditions is probably offset by the fact that this group showed a more rapid decline in performance over the three days than did any of the other groups. Social interaction appears to have the advantage of maintaining acquisition behavior at a relatively high level. While the performance of pupils on the achievement test was comparable at all three grade levels, the efficiency scores did improve with age with the older children requiring less time than the younger for each item answered correctly.

Chapter VI is a follow-up of the study which has just been considered, and involved an extension of the work to the 7th and 8th grades. In addition, it sought to determine the effect of matching or mismatching the pupils in the pairs in terms of previous academic achievement in solid subjects. The

unexpected finding appeared that the pupils who were mismatched learned the task significantly better than those that were matched. The observations of the experimenters suggest that this finding may be a result of the mismatched pupils interacting more with one another while working in pairs than did the pupils in the matched pairs. At the seventh and eighth grade the efficiency scores of the pupils were no greater than at the lower grade levels. Pupils at the junior high school level may well have prolonged the task in order to remain away from their regular rooms. In this study as in the previous one, the pupils working on the task alone showed a greater decrement over days than those working in pairs which further substantiates the position that social interaction has high value for maintaining behavior.

The data across the five grades from the 4th to the 8th shows no difference between the groups as performance was measured at the end of each week. The difference lies only in the speed with which the pupils are able to master the task. Since this conclusion was derived from a rote learning task there should be no generalization to problem-solving tasks. The latter point is made since some psychologists have recently taken the rather reckless position that all mental operations can be performed at all age levels.

Chapter VII investigates the problem of the extent to which those learning by observation, the vicarious learning group, are able to use negative and positive information. The nature of the task was such that an error on the part of the person overtly responding provided exactly as much information as a correct response. The expectation was that those learning vicariously, without overtly responding, would learn more from the correct responses of the responding subjects than from the wrong responses. The data, taken as a whole, do not give support to this hypothesis. Indeed,

no consistent trend in the data over the two studies could be found.

Chapter VIII considers a different problem from that which was considered in the earlier chapters in this section. While the data of the earlier chapters generally supports the applicability of the reinforcement model to learning simple materials in the classroom, the model would appear to have limitations. In this last chapter, a study was undertaken of the reinforcement of a class of responses which were such that one would not expect the reinforcement model to apply. Several such categories can be identified. One category of such responses would be fast and accurate responses such as must be made by the musician. Generations of music teachers have discovered that speed of performance is not achieved by reinforcing speed. On the contrary, speed is acquired through slow and accurate practice of the responses to be performed later rapidly. Speed comes automatically when the associations between the performance of successive notes and chords have been strengthened to the point where flawless accuracy of performance is achieved.

The category of responses selected for reinforcement in Chapter VIII were unusual responses of the kind that are manifested on the Guilford _Unusual Uses Test_. Although many contemporary psychologists and educators have assumed that the development of creative behavior requires only that the teacher reinforce this behavior and make it in some way rewarding, such a theory lacks plausibility. First, a careful examination of previous studies indicates that any such effect previously found may be nothing more than an experimental artifact. Second, the ability to produce unusual responses would appear to be highly dependent upon the individual's cognitive structures from which they are generated. Perhaps the problem

of facilitating the generation of unusual responses is not so much a matter of arranging for the pupil to want to make such responses as it is a matter of providing him with the necessary internally stored knowledge from which such responses are derived. One disturbing possible source of artificial results in previous experiments is that the same person was generally used as both trainer and also the administrator of the criterion tests. It seems reasonable to suppose that training may have consisted of nothing more than the student discovering what the trainer wanted when he asked for an unusual use of an object or an original response.

The study undertaken on the reinforcement of unusual responses was planned so that the devices used to determine whether reinforcement had produced an increase in the ability to produce original responses were adminstered by the teacher as a part of the regular class routine. The training tasks were administered by the experimenter and were understood by the children to be part of a study. The training task required the subjects to think of unusal uses of common objects - ten objects a day for 15 days. The main evaluation tasks were derived from the work of Torrance and involved the interpretation of colored slides.

No training effect was found, which is exactly what was expected. This result raises doubts about the validity of the work of other research workers who have not controlled some of the conditions in which an apparent training effect appears as an artifact.

Substantial correlations were found between the measures of creativity and measures of achievement. These correlations are hardly surprising. The child who obtains a high score on any of the tests of creativity included in the study is likely to display on the same test a high level of knowledge.

Indeed, the development of the child's knowledge would appear to be the key to the development of creativity as it was measured by the tests. Reinforcement would not appear to be the key to the problem in terms of the data which are presented in this report.

In addition, very little evidence could be found for the belief of some psychologists that teachers discriminate against the creative child. On the contrary, the data came nearer to supporting the position that teachers favor such children and tend to give them high grades.

i III

SECTION III

STUDIES COMPARING DIFFERENT REINFORCING CONTINGENCIES IN CHILDREN

Introductory Statement to Section III

If the reinforcement paradigm can be applied to the planning of human learning situations, then it becomes very important to know the reinforcing value of different stimuli. While there are many prejudices current in education concerning this matter, knowledge is still meager. Typical among the widely held views of persons connected with education is the view that human beings are particularly effective reinforcers of human behavior. The theory underlying this point of view is that children seek approval and that hence approval can be used to control human learning. If the young human is an approval-seeking organism then approval might well be expected to have a high reinforcing value. Such a conception of the events that are most reinforcing for human behavior has very important implications for the planning of education. Insofar as the position is sound, one would expect that direct approval by a teacher would, then, be more effective than the very indirect approval provided by a teaching machine. The importance of this problem has led the staff of the project to give consideration to the study of approval as a reinforcer in contrast to the reinforcement provided by physical events which are not directly symbolic of approval. The first chapter in this section of the report represents a study which attempts to compare approval with a physical reinforcing contingency which is not directly approval-related.

A second view commonly expressed in educational circles is that any unpleasant condition is likely to depress learning. This view which has become identified with the progressive education movement represented a reaction against the somewhat free use of aversive stimuli in

in many schools of the past. The arguments against the use of aversive stimuli are many and varied. While one is that aversive stimuli depress learning, other arguments against its use have also been advanced by educators on moral and ethical grounds. The position advanced by such arguments has been supported by psychologists, but on different grounds. Psychologists have commonly advanced as an argument against the use of aversive stimuli the fact that such stimuli elicit anxiety and that anxiety thus may become conditioned to a whole range of common stimuli. An additional argument is that there is also generalization of the anxiety response to stimuli which are similar along some dimension to that which produced the anxiety in the first place. However, such arguments are generalizations from studies involving subhuman organisms which have been exposed to very severe shock. The fact is that relatively little is known concerning the effect of less intense aversive stimuli on human learning. For this reason two studies have been conducted which compare the effects of pleasant and noxious stimuli operating within the framework of the reinforcement paradigm. These studies are reported in the second chapter of this section.

SECTION III

CHAPTER I

EFFECTS OF DIFFERENT REINFORCERS: A COMPARISON ACROSS AGE LEVELS

This study provides a comparison of the relative effectiveness of two reinforcing conditions, referred to as verbal and physical, as they effect responses made by boys and girls to a knob-pressing task. The hypotheses state that there are differences in the reinforcing value of the two reinforcing events which are related to the age and sex of the subjects.

The two classes of reinforcing events considered in the study are approving statements made by the experimenter (E), which are termed verbal reinforcements, and the delivery of a marble by the apparatus, termed physical reinforcement.

While the study limits itself to the two specific reinforcing contingencies described, the problem is of interest because of a dearth of information concerning the relative reinforcing value of different events. Some of the problems related to the use of teaching machines cannot be solved until much more is known about the reinforcing value of different classes of events with different age groups.

The study is concerned with two reinforcing events only. However, the ultimate concern of this and similar studies is to identify classes of events which have particular value for reinforcing human behavior.

Some General Conditions Involved in Reinforcement Studies with Children

The concept of reinforcement implies a stimulus event which will maintain or increase the strength or frequency of a response. Most reinforcers that are effective in the shaping of human behavior are only very remotely related to primary reinforcers. These reinforcers

are called secondary reinforcements and they derive their power from the fact that each symbol that reinforces represents a class of situations which in themselves have highly reinforcing properties. Such symbols as are used as reinforcers may represent different classes of events for different persons. The child who solves a problem and is reinforced by the teacher's announcement of "good," may interpret this announcement as a sign of personal achievement or as a sign of having achieved status in the group.

Money, grades, toys and marbles have served in previous studies as physical or material symbols which have reinforcing properties and which hence increase frequencies of response. Saying "good" or "that's good" are obviously successful as verbal symbols of approval in both verbal and motor tasks.

Reinforcers are related to behavior by certain laws as are other aspects of the environment. Meehl (1950), for example, states that most reinforcers for operant behavior are trans-situational. That is, if an object or event has proved successful in strengthening responses in an organism in one situation, it will strengthen all or most learnable responses in the same organism. The term operant behavior is used here merely to distinguish this type of learning or responding from that involved in classical conditioning.

In terms of the trans-situational law proposed by Meehl then, any of the preceding reinforcers should be relatively effective in bringing about an increase in response strength where a learnable response is employed.

The criterion of the effectiveness of a reinforcement is a change

in the strength of a response as a result of the reinforcing stimulus event. One method of quantifying the construct of strength and the one which is used in this study is through procuring a measure of the frequency of response.

In order to measure and compare different reinforcers across age levels, the same type of device must be utilized with all age groups. The device must surmount the barrier imposed by differences in language and reading skills and must afford a basis for comparison of responses. Manipulative or mechanical apparatus have typically been selected to meet the imposed criteria.

Terrell (1958), in comparing performance at two age levels, used a "greater than" concept task where the subject was to choose the larger or smaller of two similarly shaped objects, i.e., between two cubes, squares or triangles placed before him.

Another commonly used device involves an apparatus where knob-pressing results in the delivery of one or more types of reinforcement. These devices are generally variations on a theme and differ with the design of the study and the desires of the experimenter. Jones and Liverant (1960), for example, designed the exterior of their two-knob device with a clown's face from whose mouth reinforcements could be dispensed. McCullers and Stevenson (1960) used a device which had three centrally-located knobs, designated to deliver different kinds or combinations of kinds of reinforcements to which the subject could respond. A device of the latter type affords a basis for comparison of responses to different stimuli and can likewise be utilized across different age levels without an alteration of the task.

Incentive Items

Studies of reinforcement in children commonly provide the child with a small gift which he knows, in advance, he is to receive if he performs well on the task. For example, McCullers, *et al*. (1960) and also Kuenne (1956) provided small toys which each child was told he was to receive and which were visible while the task was being performed. In some situations, such an over-all incentive item is unnecessary, as in the study of Alberts and Ehrenfruend (1956) in which gumdrops were used as the reinforcing event.

The incentive items used are generally those towards which children of the particular age group used show a stong positive response. Such items themselves would almost certainly function as reinforcers, but their function in the particular setting is to maintain an orientation towards the task. Toys and candy have been well-established as incentive items.

A range of such incentive items from which the child chooses seems to be a good arrangement. Brackbill and Jack (1958) demonstrated that individual preferences rather than sex or age *per se* determine the choices of children for such rewards. When these investigators allowed five-year old boys to choose marbles or candy or trinkets, each one of the three was chosen an equal number of times.

Results of Previous Studies

McCullers *et al*. (1960) utilized the knob-pressing apparatus previously described to yield a comparison between subjects three to five years of age and persons eight to ten years of age. A marble reinforcement

was available to the subjects upon the depression of any one of the three knobs. Coupled with the latter physical reinforcement on one knob was a verbal comment denoting approval, such as, "That's good." The effectiveness of verbal reinforcement would be demonstrated by an increase in frequency of response to the knob delivering both physical and verbal reinforcement. Their hypothesis that the younger group would be reinforced to a greater degree by verbal reinforcements than would the older group was confirmed (p = .01). The older group distributed their responses more uniformly across the differently reinforced responses and a t test revealed no significant increase in response rate to the knob that delivered a verbal reinforcement as well as the marble which all knobs delivered.

Jones _et al_. (1960) investigated another factor, namely, the ability to discriminate among the percentages of rewards available. Their study supports the preceding findings of McCullers _et al_. (1960) and offers supplementary information concerning the variability of response phenomena. Their subjects were aged four to six and nine to eleven. An apparatus with two knobs was employed. The percentage of reinforcement on one side exceeded that on the other. Two schedules were established. One schedule dispensed 90 per cent reinforcement on one side and 10 per cent on the other; the second allowed for 70 and 30 per cent reinforcement.

It was hypothesized that the younger group would respond almost 100 per cent to the side delivering the larger percentage of reinforcement while the older group would distribute their responses to both sides more nearly in accord with the percentages of reinforcement available. If the distribution of reinforcements to the two knobs were

reversed, the older children would be expected to discriminate the change sooner and respond accordingly.

Both hypotheses were confirmed which indicates that in children nine to eleven, an ability to discriminate between percentages of available reinforcements affects the responses which they make to each possible choice.

A study by Stevenson and Cruse (1961) compared responses made to a marble-insertion task by five and twelve year olds under four conditions. In one condition, the experimenter was present and attentive to the subject but made no comments; in another, the experimenter was in the same room but was not in close proximity to the subject nor did he communicate with him; in a third instance, comments denoting approval were delivered; a fourth condition involved the delivery of critical comments by the experimenter.

The findings showed that the twelve-year-old group performed at a lower level than did the five-year old group under conditions of reinforcing comments and attention from the experimenter. Stevenson et al. (1961) interpreted these findings as indicating less dependence on adult approval among twelve-year olds than among five-year olds.

Page (1958) utilized the comparative method in exploring the relative effectiveness of three methods of distributing grades and the reinforcing contingencies involved. He employed the performance of two groups composed of males and females...one of junior high school age and the other of senior high school age.

A letter grade issued with no comment, a letter grade issued with a specified teacher comment, and a letter grade issued with an appropriate

and spontaneous teacher comment (called a "free" teacher comment), comprised three reinforcing conditions which were being compared. Following the random assignment of these reinforcing conditions to the three experimental groups, the students received the grades which they had made on a specific test. On the next test administered to the groups, the test scores obtained by the students in the groups wherein grades had been distributed with specified and free comments were significantly higher ($p<.01$) than those obtained by the students in the no-comment group. A trend toward higher scores in the free comment group than in the specified comment group was shown but the difference was not statistically significant. Page states that the findings indicate a greater responsiveness to adult approval among the high school students than among the subjects of junior high age though the differences were not statistically significant.

The studies of McCullers _et al._ (1960) and Stevenson _et al._ (1961) demonstrate that at ages three to five, adult approval may be sought and will serve as an effective reinforcer; at ages nine to twelve, adult approval results in no greater frequency of response than do physical reinforcements and therefore is not considered a particularly effective reinforcer at this age level.

Page's study (1958) indicates a tendency for adult approval to be more effective with subjects of high school age than with those of junior high age. Specific age levels were not defined by Page and in generalizing from his findings, some assumptions were necessarily made concerning the ages involved.

Interaction effects based on sex differences between the subjects and experimenters have been investigated at some age levels, particularly among pre-kindergarten and kindergarten students. The findings have been inconclusive but this dimension may have some relevance for this study.

In a situation where the task was easel-painting, Gewirtz (1954) obtained a measure of the degree of responsiveness to the experimenters through the frequency of attention-seeking behaviors and comments made by the subjects in the presence of the experimenter. With male subjects aged four to five and one-half, there was a significantly greater frequency of these responses made in the presence of the female experimenter than in the presence of the male experimenter. Female subjects likewise demonstrated a greater frequency of response in the presence of an experimenter of the opposite sex than with an experimenter of the same sex though the difference was not statistically significant.

Gewirtz (1954) postulated that the interaction he found might not obtain at other age levels and explained his findings in terms of Freud's theory regarding development during the phallic period and particularly at the Oedipal stage. In its simplest form, the Oedipus complex implies a positive drive toward the opposite-sex parent with accompanying feelings of hostility toward the parent of the same sex. It is theorized that boys maintain the mother or female as an object to be chosen over the father or male until the latter part of the fifth year; girls transfer their choice preference from the mother to the father around age four.

Utilizing kindergarten subjects (ages five years, three months to five years, seven months), Rosenblith (1959) found more learning occurred with male experimenters than with female experimenters with both male and female subjects. The difference was statistically significant only between the female subjects and the male experimenter. In this study, the Porteus-maze task was learned by imitating the leader.

The findings of previous studies support the following conclusions:

1. On a simple knob-pressing task such as was used by McCullers et al. (1960) there is a tendency for approval to increase frequency of response in the three to five year olds but not in older children. However, older children respond to other features of the task which the younger do not --a point brought out by Jones et al. (1960). Some further evidence of the effectiveness of adult approval in the case of the pre-school group was found in the Stevenson et al. (1961) study. The latter study used a marble insertion task rather than a knob-pressing task.

2. Not all evidence suggests that approval is a poor reinforcer for children in the upper elementary grades. Page (1958) found evidence that a written comment of approval appeared to function as an efficient reinforcer for children, a little older, at the junior high school level. Perhaps a spoken comment of approval is less effective than a written comment because of the transitory nature of the spoken word. Approval may become increasingly effective through the teens, a matter requiring further study.

3. The findings suggest the existence of interactions between the sex of the experimenter and the sex of the subject in the effectiveness

of approval as a reinforcer. Gewirtz (1954) indicates that his data supports the contention that male subjects are more readily reinforced by an experimenter of the opposite sex at ages of four to five and a half. Girls aged five years, three months, to five years, seven months were also more readily reinforced by an experimenter of the opposite sex.

The Present Study

Previous studies have not provided a direct comparison of the reinforcing value of statements of approval in contrast to other kinds of events. The purpose of the present study is to determine the relative reinforcing value of orally administered statements of approval in contrast with the reinforcing value of a physical event consisting of the delivery of a marble. In a sense, the delivery of the marble involves more than just the mechanical event, for the subject is also provided with the information that the delivery of a marble indicates, in some way, that his behavior is correct. In other studies where the effect of verbal approval has been introduced as a factor, such approval has occurred simultaneously with some other reinforcing event such as the delivery of a marble.

The assessment of the reinforcing value of comments of approval in comparison with other mechanically administered reinforcing events provides knowledge of considerable importance for the design of classroom learning situations. Traditional teaching has depended largely on approval administered by the teacher as the main reinforcing event. The difficulties of administering such reinforcements on a sufficiently

frequent basis has been responsible for attempts to design machines which provide reinforcing contingencies. Such machines depend for their efficiency on some reinforcing event other than that of spoken approval. The reinforcing events provided by mechanical devices do indirectly symbolize approval of the teacher, much as the delivery of a marble in the present experiment symbolizes indirectly the approval of the experimenter.

The study involves the testing of the following hypotheses:

1. Verbal reinforcement is more reinforcing than physical reinforcement in pre-kindergarten children.
2. Physical and verbal reinforcements are equally effective for third grade children.
3. A low level of dependence on adult approval is reflected in the sixth grade and junior high-school by the greater effectiveness of physical than verbal reinforcement.
4. Verbal reinforcement is more effective at the pre-kindergarten than the third grade level.
5. Approval is more reinforcing among females in kindergarten and third grade than among males at the same age levels.
6. Teacher's rankings on need for approval correlate with responsiveness to verbal reinforcement.

Method

Subjects

The Ss were 100 males and 100 females selected on the basis of chronological age and sex. Twenty males and 20 females were selected

within each of the following five age ranges: (a) 3-11 to 5-8, (b) 5-9 to 6-6, (c) 8-11 to 9-7, (d) 12-5 to 13-3, and (e) 14-8 to 15-8. The age levels are representative of students in pre-kindergarten, kindergarten, and grades three, six and nine. The Ss in the pre-kindergarten group were obtained from private nursery schools and the remainder from two public schools in metropolitan Salt Lake City.[1]

Apparatus

A device which delivered different kinds of reinforcement upon the depression of each of four knobs was selected as a task which could be used at the various age levels (see Fig. 1.01). The turquoise panel which fronted the apparatus was a square, 23 1/2 inches across. Centered 8 1/4 inches from the top edge was a round, red light which served as a signal to S for another trial. The light was turned on from behind the panel and extinguished automatically following the depression of a knob. The aluminum knobs were centrally placed 7 1/2 inches below the red light. A glass-covered, enclosed box located 3 1/2 inches below the knobs served as a receptacle for the marbles which were dispensed upon the activation of a solenoid behind the panel. The contents of the receptacle were not accessible to S but entry of the marbles into the box was audible and the marbles were visible.

Procedure

The Ss were tested individually in an experimental room established in the school plant. The E addressed each S by his first name and gave these instructions:

[1]The writers express appreciation to the faculties of the schools involved and to the students involved for their cooperation.

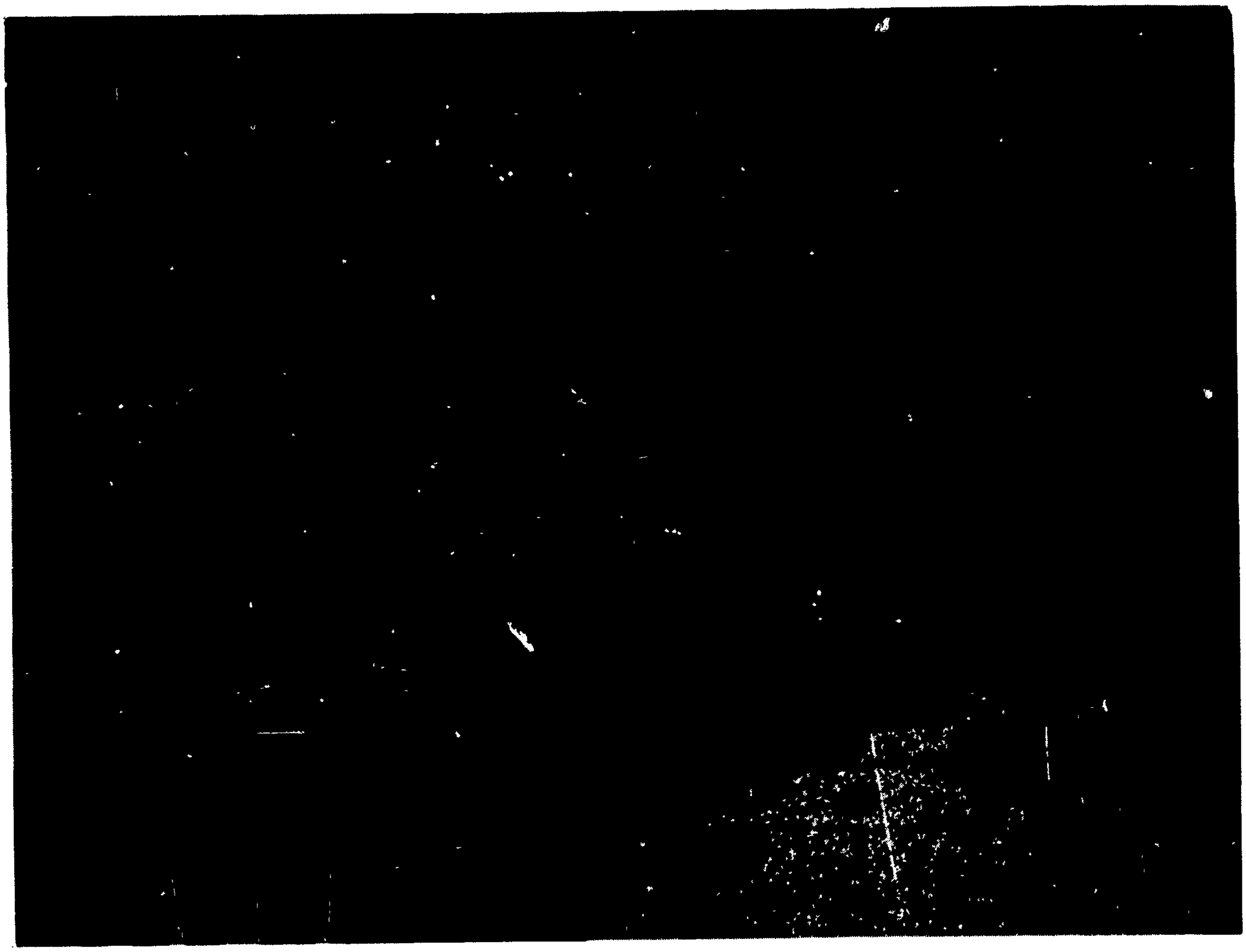


Fig. 1.01 Arrangement of apparatus with experimenter at the left and the subject at right.

"__________, this is a game. I can turn a light on like this (light on), and you can turn it off (depress knob). Now, you try it (light on). Now, try another one (light on). Try all of them. (Allow a total of four trials.) <u>But</u> not all of the knobs are the right knob. Sometimes I will tell you that you have pushed the right knob. I'll say 'good' or 'that's good.' Sometimes you'll get a marble here (pointing) when you push the right knob. If you push the right knob most of the time, you may choose one of these toys (point to display) when we're through. Now, how will you know if you've pushed the right knob? (Continue with this until they say 'get a marble' and 'you say good.')"

Practice was given under actual reinforcement conditions until S had depressed each of the four knobs one time. Following this, S was given 120 trials.

The conditions assigned to the four knob positions were marble only, verbal, marble and verbal simultaneously, and no reinforcement. To counterbalance knob position effects, the conditions were assigned across the knob positions so that each reinforcement was delivered from each knob position an equal number of times within each age group. Reinforcement was delivered on 50 per cent of the trials and was assigned so that it occurred in random, rather than systematic order.

Incentive items were provided for participation and were displayed approximately six feet from S. A bag of candy, a toy suitable for a male, one suitable for a female, and another which might be chosen by either sex constituted the four-item display. Objects which were appropriate to the age of the S, were selected at the four age levels. The objects were held constant within each age group.

Results

In order to function as reinforcers, the reinforcing stimulus events must maintain or increase the frequency of the response upon which it is contingent. Figure 1.02 shows the mean number of responses made to the various stimuli by blocks of twenty trials for the total sample.

The responses made at various age levels to the four treatments are presented for comparative purposes in Figure 1.03. Figures 1.02 and 1.03 demonstrate:

1. Response in the absence of the two reinforcers provided does not extinguish even for this relatively long task.

2. Response in the absence of the two reinforcers provided decreases as a function of age with the greatest frequency of response in the pre-kindergarten group and the least in the junior-high group.

3. In the total sample, the frequency of response in the presence of approval reinforcement decreases in magnitude in the last block of twenty trials from the frequency observed in the first block of twenty trials. The difference is not statistically significant ($\underline{t}$ = 0.74).

4. A decrease in response to approval reinforcement across the blocks of trials is noted in the kindergarten and third grade group; in the sixth grade and junior-high groups, the frequency of response is maintained at approximately the same level when comparing the first block of trials with the last; in the pre-kindergarten group, an increase in response is noted.

5. The mean number of responses in the presence of marble-verbal and marble reinforcement are typically greater in magnitude than are

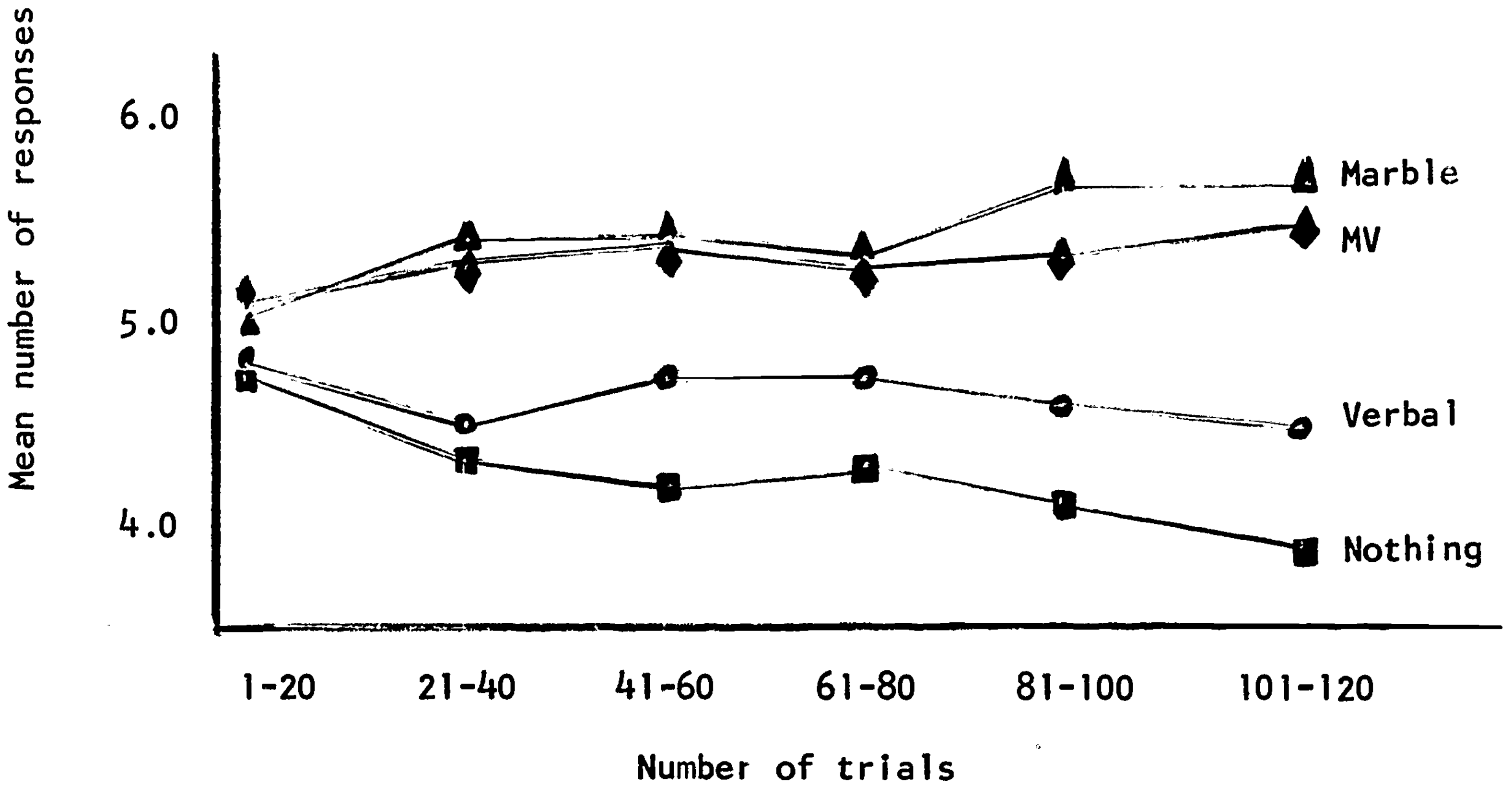


Figure 1.02 Mean number of responses for all male and female subjects for each block of 20 trials.

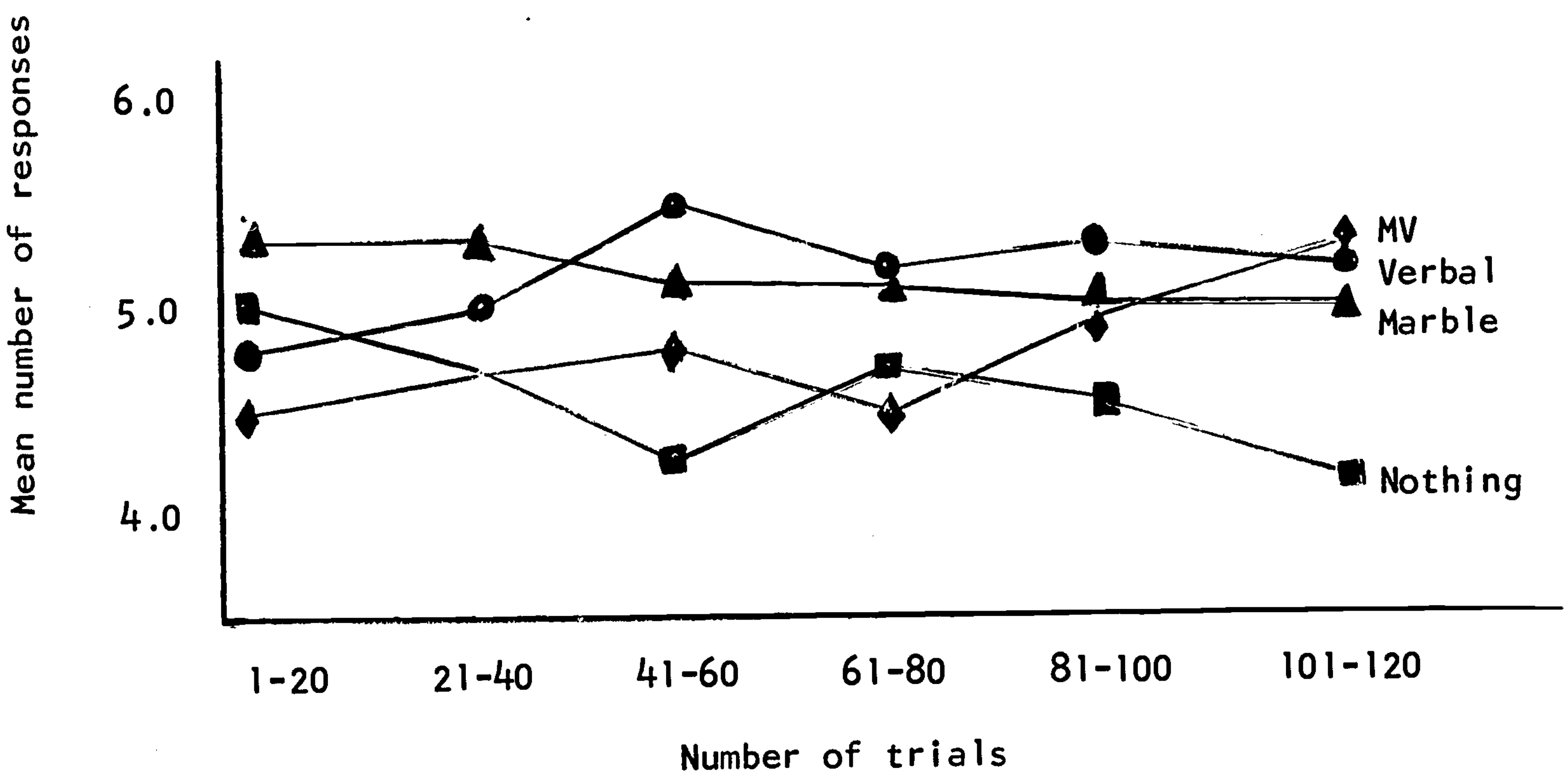


Figure 1.03a Mean number of responses for male and female subjects in the Prekindergarten group.

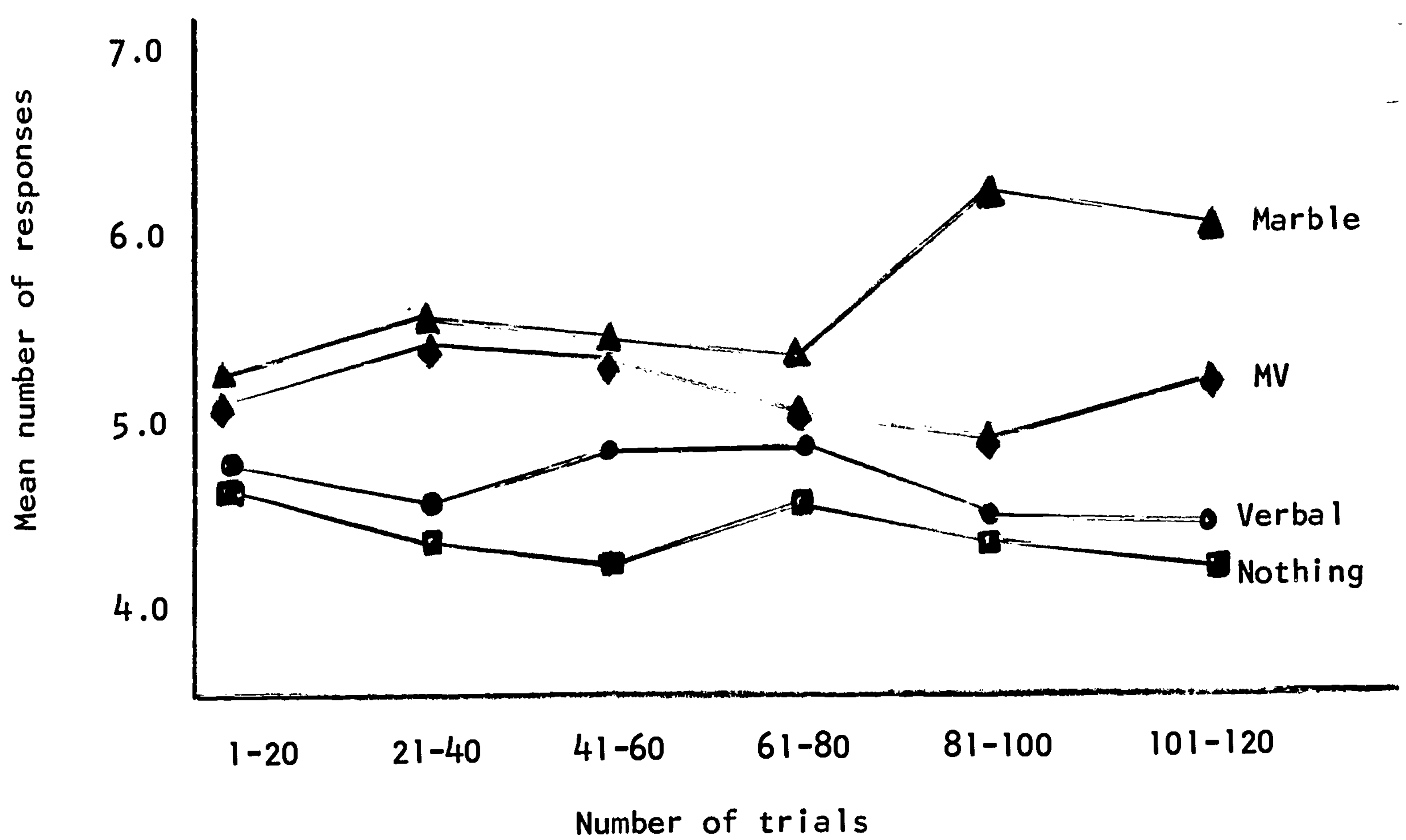


Figure 1.03b Mean number of responses for male and female subjects in the Kindergarten group.

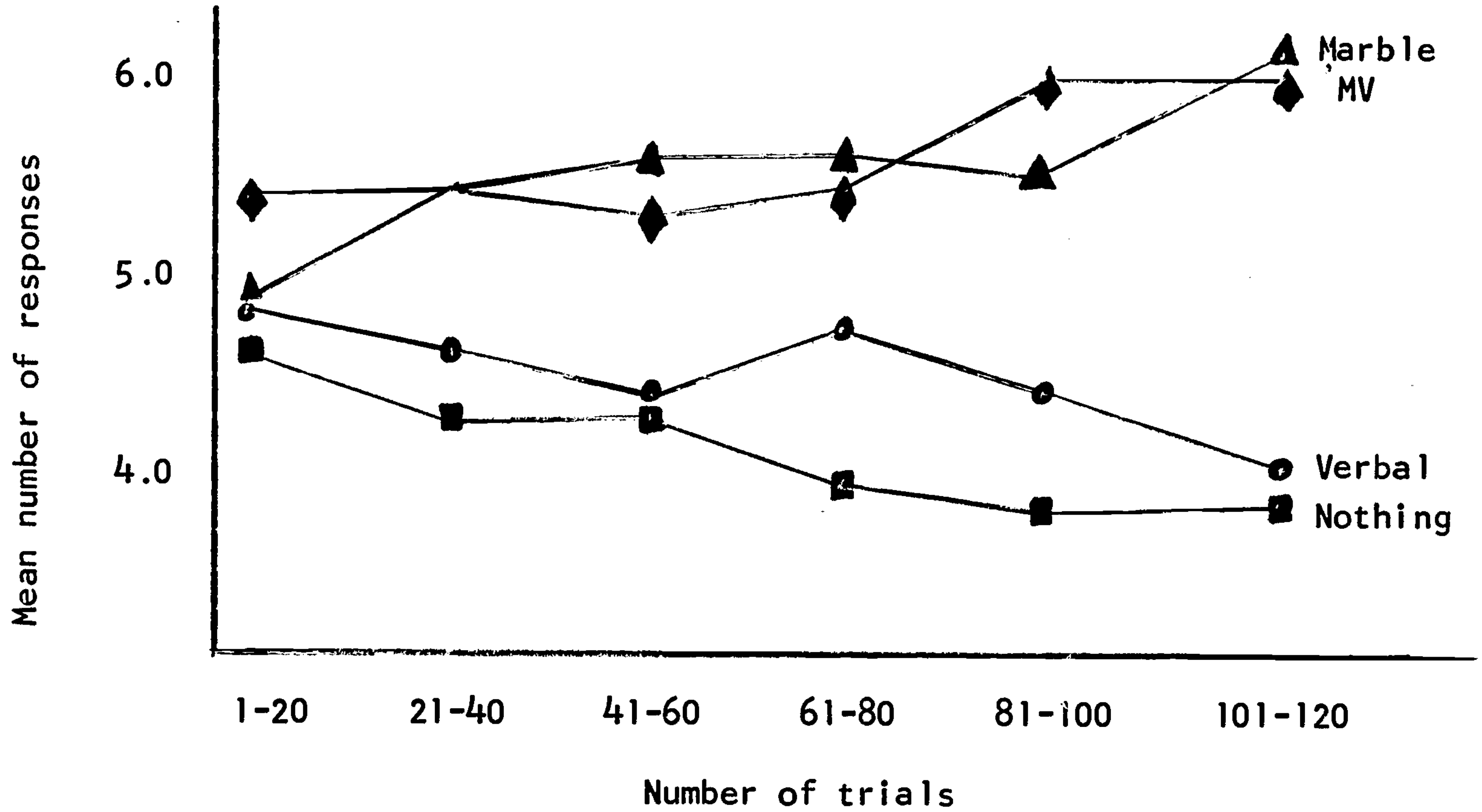


Figure 1.03c Mean number of responses for male and female subjects at age level 8-11 to 9-7 (third grade group).

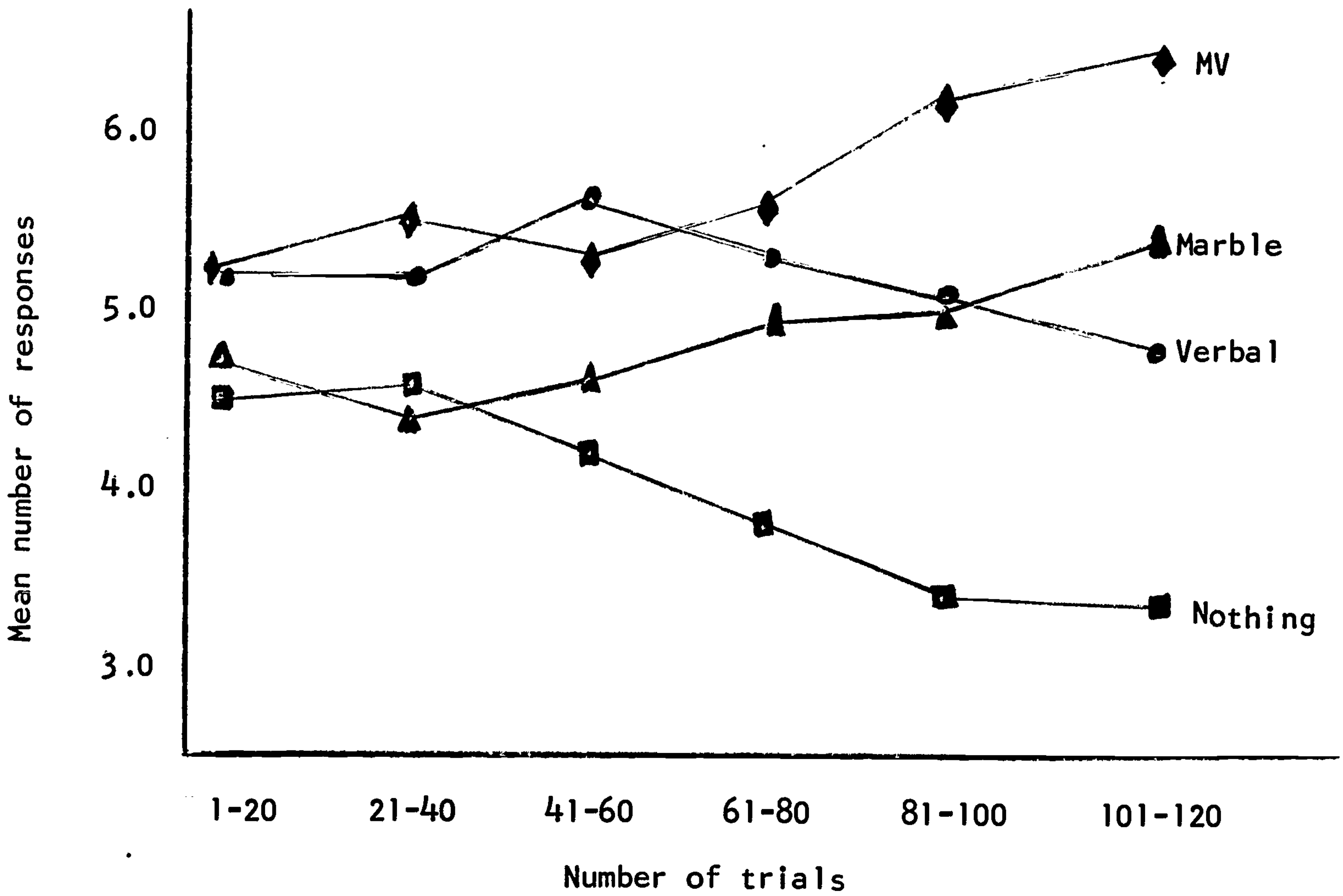


Figure 1.03d Mean number of responses for male and female subjects at age level 12-5 to 13-3. (sixth grade group).

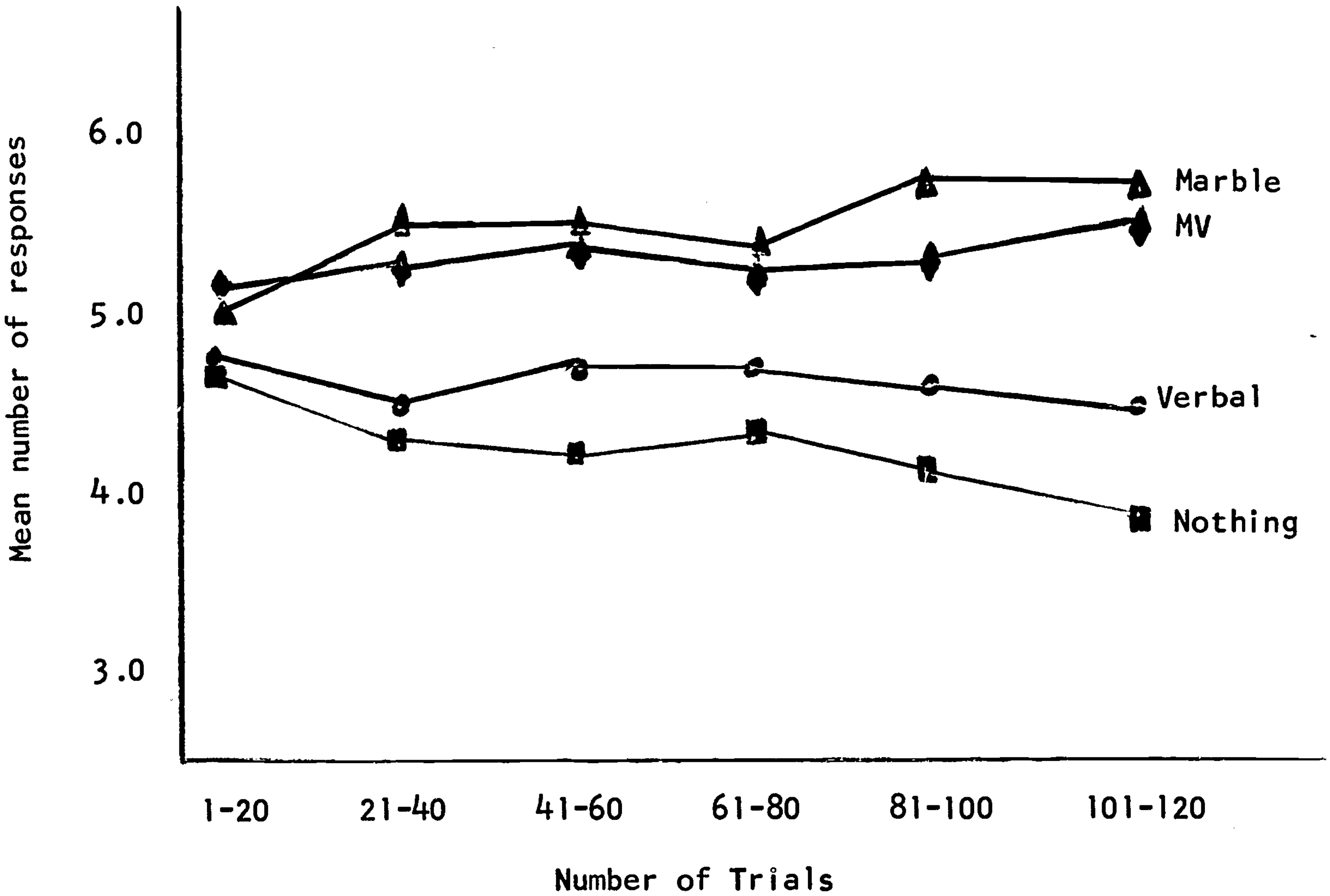


Figure 1.03e Mean number of responses for male and female subjects at age level 14-8 to 15-8. (ninth grade group).

the mean number of responses made in the presence of verbal stimuli. The only exception is found in the pre-kindergarten group.

6. The greatest amount of learning is shown by Ss in the two oldest groups where fewer responses occur in the absence of reinforcement.

The learning curves of the five groups provide very clear evidence that physical reinforcement is much more effective than the verbal. Table 1.02 provides an F-test to determine the significance of the difference between reinforcing contingencies. The difference between the effects of the reinforcing conditions have a very high degree of confidence attached to them. It should be noted here that there are difficulties involved in undertaking analyses of variance on the complete data since restrictions are introduced by the fact that the total number of responses is always 120. This restriction becomes less troublesome when the data from the two main reinforcing conditions are involved as is the case in subsequent analyses of variance.

The basic data on which the following analyses of variance are based is given in Table 1.01. Table 1.03 presents a 2 x 2 x 5 analysis of variance involving the two major reinforcing conditions, two sexes, and five age groups. A highly significant main effect related to reinforcing conditions is evident. An interaction between reinforcing conditions and sex is significant at almost exactly the five percent level. This latter effect appears to be mainly due to two factors: (a) the responses of the pre-kindergarten children to the task and (b) the responses of males to physical reinforcement. The differences between boys and girls at the pre-kindergarten level are shown in the learning curves presented

Table 1.01

FREQUENCY OF RESPONSE TO EACH STIMULUS

FOR 20 BOYS AND 20 GIRLS AT EACH AGE LEVEL

	Verbal		Marble		Marble-Verbal		Nothing	
Pre-Kindergarten								
Male	546		653		627		574	
Female	621		603		633		543	
Sub Total		1167		1256		1260		1117
Kindergarten								
Male	522		708		626		544	
Female	602		651		623		524	
Sub Total		1124		1359		1249		1068
Third Grade								
Male	563		684		662		491	
Female	536		653		689		522	
Sub Total		1099		1337		1351		1013
Sixth Grade								
Male	548		638		709		505	
Female	605		647		675		473	
Sub Total		1153		1285		1384		978
Ninth Grade								
Male	548		656		703		493	
Female	579		632		681		508	
Sub Total		1127		1288		1384		1001
Totals								
Male	2727		3339		3327		2607	
Female	2943		3186		3301		2570	
Totals both sexes		5670		6525		6628		5177

Table 1.02

ANALYSIS OF VARIANCE FOR TOTAL SAMPLE

	df	Ms	F	p
Between Reinforcing Conditions	3	2427.06	42.78*	<.001
Within	796	56.73		
Total	799			

*Significant beyond .001 level.

Table 1.03

ANALYSIS OF VARIANCE OF VERBAL AND PHYSICAL REINFORCING CONDITIONS

	df	Ms	F	p
Reinforcement	1	1827.56	34.68	<.001
Ages	4	8.68		
Sex	1	9.925		
R x A	4	52.91		
R x S	1	340.41	6.46	.01
A x S	4	25.40		
R x A x S	4	39.08		
Within	380	52.69		
Total	399			

in Figure 1.04. The girls show a clearly greater responsiveness to verbal reinforcement than do the boys. Males respond with higher frequency to physical than to verbal reinforcement at all age levels.

A 2 x 3 x 5 analysis of variance (Table 1.04) was performed including three of the reinforcing conditions...verbal, physical, and physical-verbal. There was, again, a difference among the conditions of reinforcement significant beyond the .01 level. A reinforcement by sex interaction was significant at approximately the .05 level (_F_ obtained = 2.89; _F_ required = 2.99). Scheffe's test among the means of the reinforcing conditions in this analysis demonstrated that physical reinforcement alone was significantly more effective than verbal alone. A comparison of the mean of the physical reinforcement with the mean of the verbal-physical combination was also performed in an effort to determine the effect of providing verbal and physical reinforcement simultaneously. No significant difference between the two indicates that, statistically, physical reinforcement alone is as effective as physical reinforcement when it is offered simultaneously with verbal reinforcement.

At the lower age levels significant differences were found between the number of responses made by males and females to the knob reinforced with approval. The differences in the mean number of responses for males and females for the pre-kindergarten and kindergarten age group are shown in Table 1.05. Consistent with previous work, the significant differences are at the kindergarten and pre-kindergarten levels. The learning curves from which these differences are derived are shown in Figure 1.04.

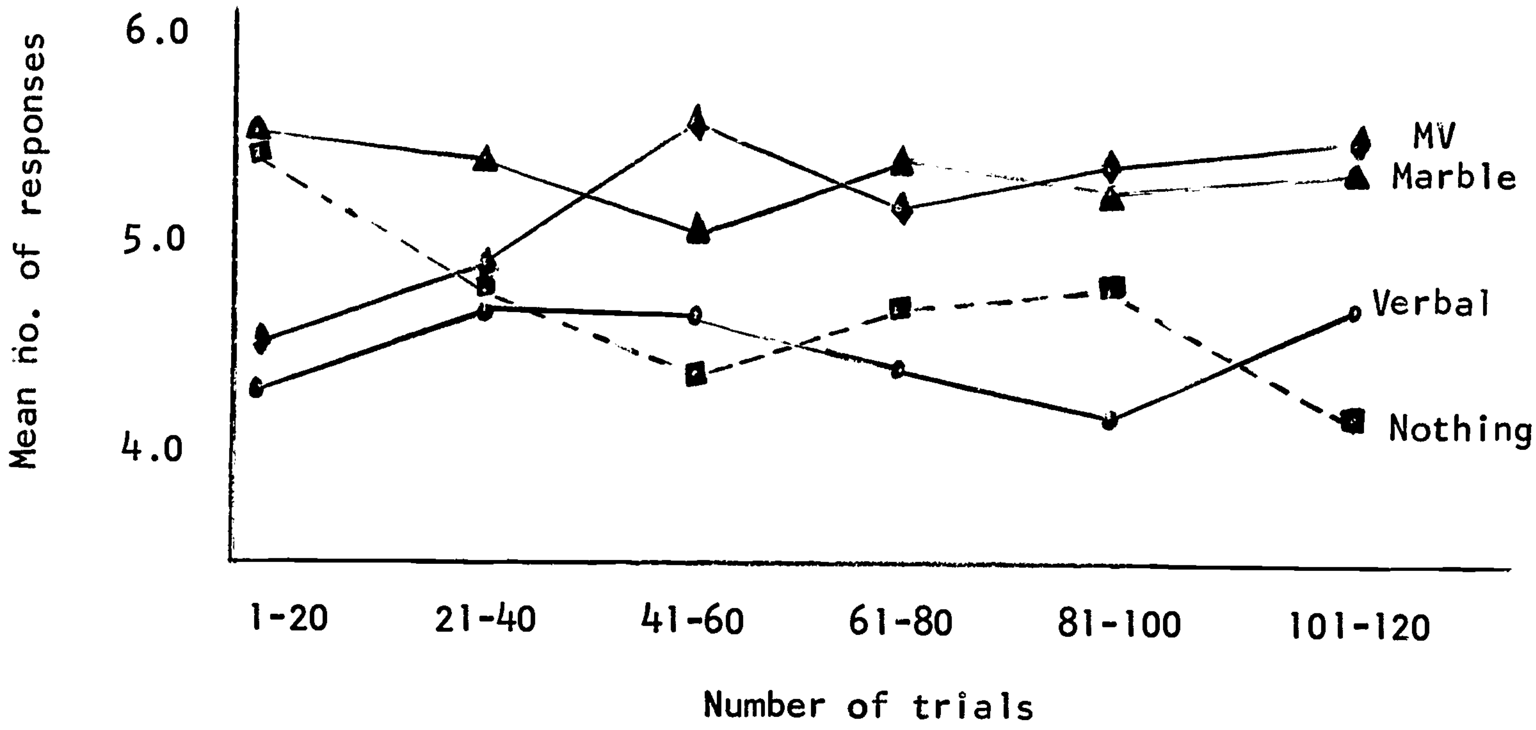


Figure 1.04a Total mean responses of male subjects (N=20). Pre-kindergarten group.

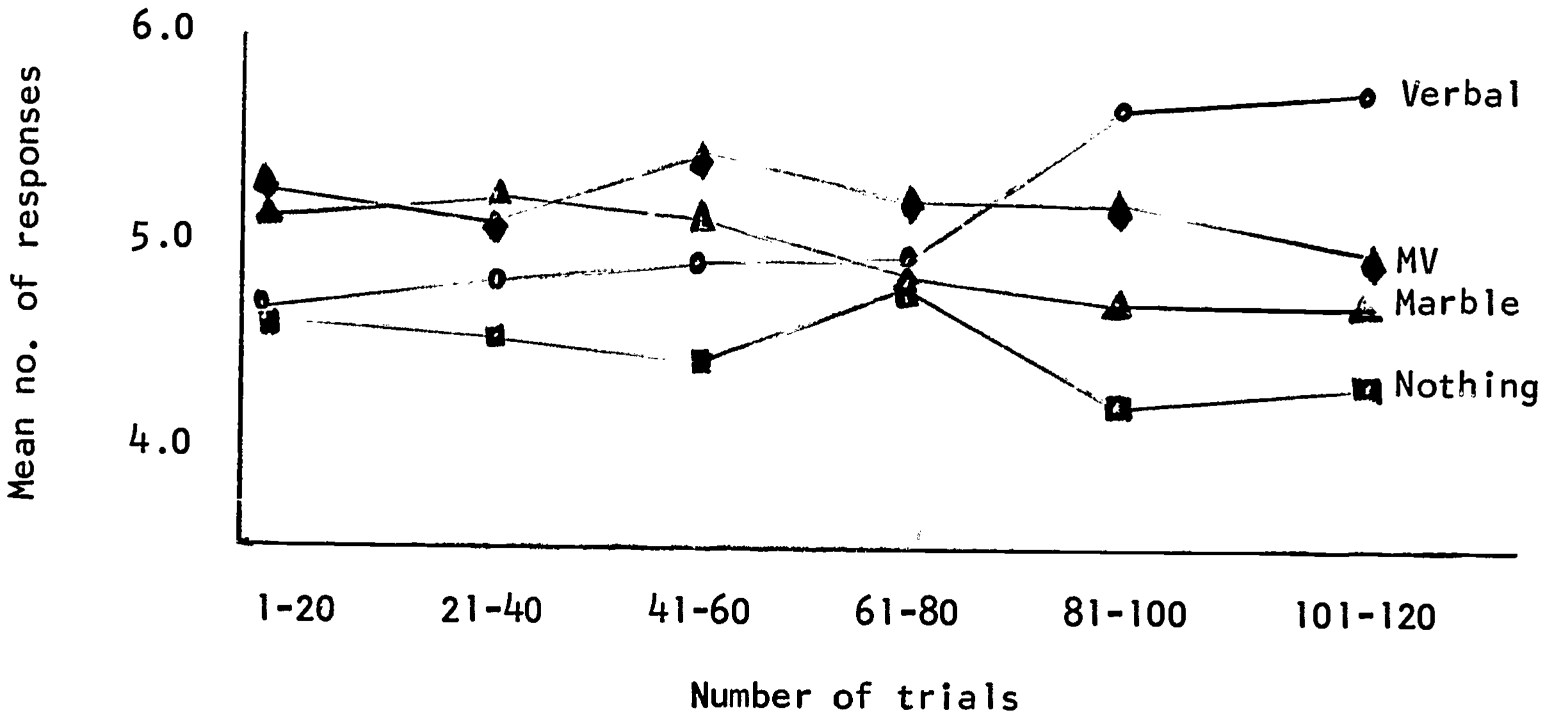


Figure 1.04b Total mean responses of female subjects (N=20). Pre-kindergarten group.

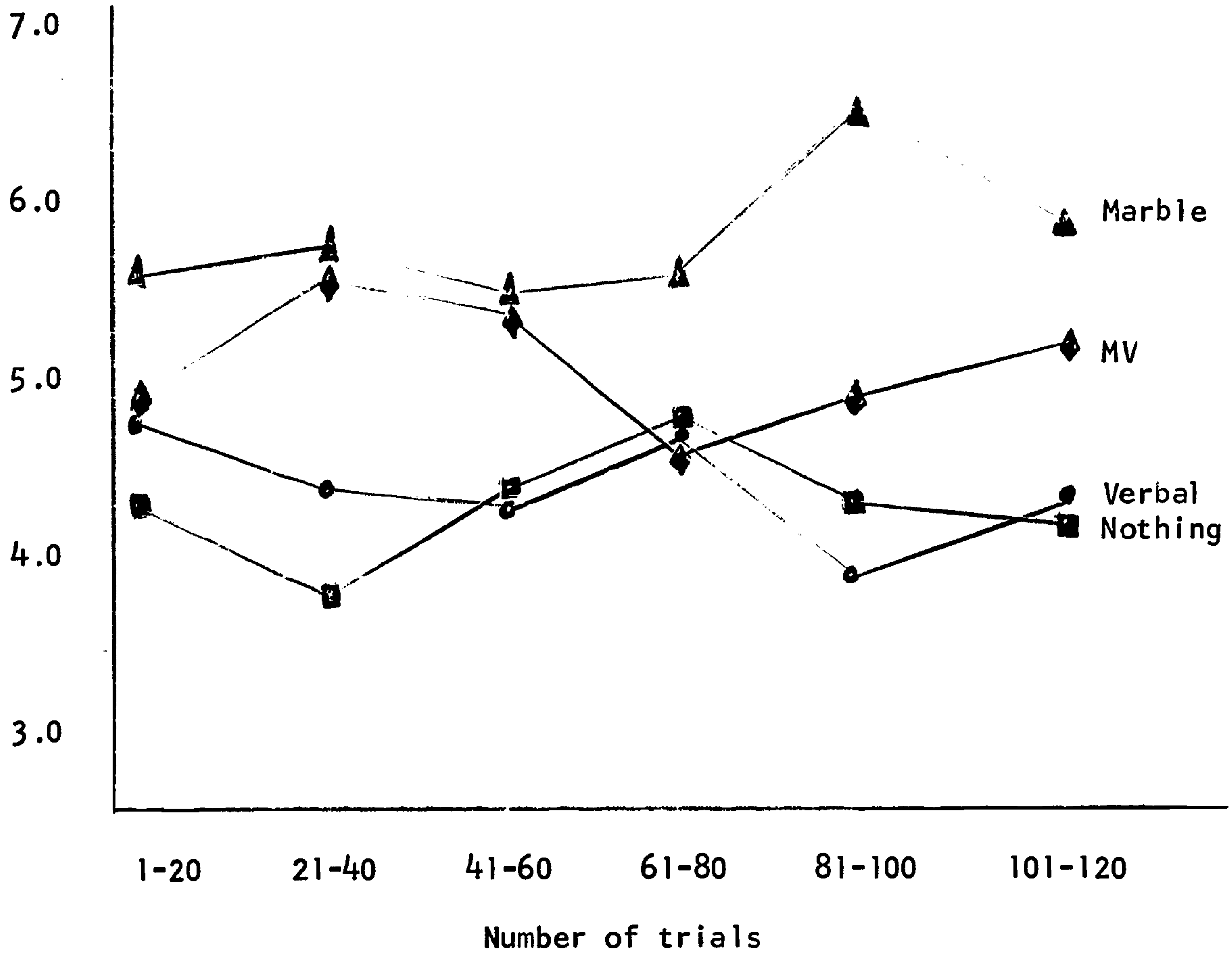


Figure 1.04c Mean responses of male subjects in the Kindergarten group.

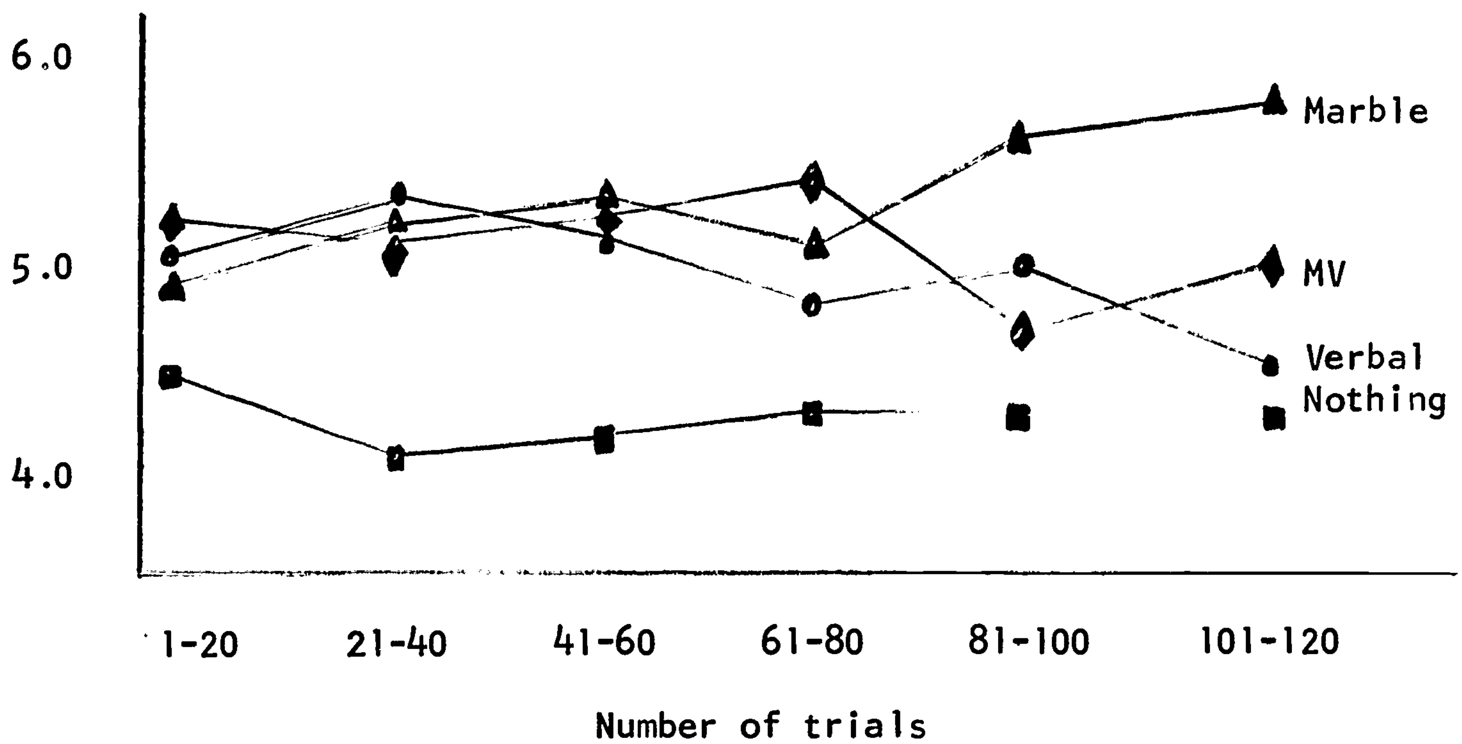


Figure 1.04d Mean responses of female subjects in the Kindergarten group.

Table 1.04

ANALYSIS OF VARIANCE OF VERBAL, PHYSICAL AND PHYSICAL-VERBAL REINFORCING CONDITIONS

	df	Ms	F	p
Reinforcement	2	1382.83	22.58*	<.001
Ages	4	26.46		
Sex	1	2.28		
R x A	8	72.68	1.19	
R x S	2	175.72	2.87**	.05
A x S	4	6.87		
R x A x S	8	35.93		
Within	570	61.24		
Total	599			

*Significant beyond .001 level.
**Need 2.99 at .05 level.

The latter finding does not necessarily imply that responses to the verbally reinforced knob are more frequent at the lower than at the higher age levels. In actual fact there is no significant difference between the mean number of responses to the verbally reinforced knob of the pre-kindergarten and the third grade group ($t = 1.33$). However, when the same comparison is made using only females, a difference is found which is significant at the .01 level of confidence, as is shown in Table 1.06.

Rated Need for Approval of Pupils and Responsiveness to Verbal Reinforcement

The teachers of the school children included in the study were asked to rank the children in their classes for their need for approval. The directions given to the teachers in making the rankings were as follows:

> "Please rank the individuals below according to the way they respond to approval. There are ____names included in the list. Number 1 will signify the individual whom you feel is most responsive to approval or praise when compared with the other names on the list. The person whom you feel is next most responsive to approval will be number 2, and so on down the list. Number _____ will signify the individual among the group whom you feel is least responsive to praise.

The rankings thus obtained were converted to normalized standard scores. All rankings were then pooled and a correlation computed between ranking and frequency of response to the knob reinforced with verbal approval. A Pearson _r_ of .005 indicates a zero over-all relationship between the two variables. However, such a pooling of data is not

Table 1.05

MEAN FREQUENCY OF RESPONSE OF MALES AND FEMALES TO THE VERBALLY REINFORCED KNOB AT ALL AGE LEVELS

Age Group	Mean Frequency of Response: Male	Female	t
Pre-Kindergarten	27.3	31.05	2.03*
Kindergarten	26.1	30.1	2.78
Third Grade	28.15	26.8	1.07
Sixth Grade	27.4	30.25	1.79
Ninth Grade	27.4	28.95	1.15

*Requires t of 2.086 at .05 level.

Table 1.06

COMPARISON OF FREQUENCY OF RESPONSE TO VERBALLY REINFORCED KNOB FOR PRE-KINDERGARTEN AND THIRD GRADE CHILDREN

	Pre-Kindergarten Mean	Third Grade Mean	t
Female	31.05	26.8	2.84*
Male	27.3	28.15	0.43

*p = .01.

entirely reasonable for pupils were drawn all the way from kindergarten to the ninth grade.

Additional information on the rankings is provided by age level in Table 1.07. Spearman rank-order correlations were computed, using the rankings made by the eleven teachers without converting them to standard scores. An inspection of the table reveals negative correlations in the group of kindergarten subjects. The only significant correlation occurs in the junior-high group and it is accompanied by another correlation in that group which is high and positive.

Computing the correlations in this manner, even though N's are small, provides an opportunity for observation across age levels. In the group of kindergarten subjects, for example, the three correlations are negative, indicating perhaps the difficulty involved in discerning the student's need for approval or identifying those behaviors which suggest need for approval. In the group of junior-high students, however, patterns of behavior appear to be less random and more established enabling the teacher to make better judgments concerning actions showing need for approval. This is hinted by the correlations which are, as a group, higher than those in other groups. One of the correlations is significant beyond the .05 level, while another is relatively high and positive. It is of interest that these two rankings were made by the same teacher.

The main findings may be summarized as follows:

1. What has been termed here physical reinforcement is more effective at all age levels than what has been termed verbal reinforcement. The general appearance of the learning curves indicates that at the lower age level the response to physical plus verbal reinforcements is almost the same as the response to physical alone. At the higher

Table 1.07

SPEARMAN RANK-ORDER CORRELATIONS BETWEEN TEACHER'S RANKING ON NEED FOR APPROVAL AND RESPONSE TO VERBAL STIMULI

Group	Rho
Kindergarten	-.036
Kindergarten	-.518
Kindergarten	-.352
Third Grade	.334
Third Grade	.085
Third Grade	-.10
Sixth Grade	.345
Sixth Grade	-.543
Sixth Grade	-.026
Sixth Grade	.045
Ninth Grade	-.096
Ninth Grade	.417
Ninth Grade	.112
Ninth Grade	.694*

*Significant at .05 level.

age levels, there is the suggestion of an additive effect of the two reinforcing modes, however, there is no significant difference.

2. Verbal reinforcement is more effective with females than males at both the pre-kindergarten and kindergarten age levels.

3. In comparing the responses of pre-kindergarten Ss with those in third grade group, the females showed a significantly greater number of responses to verbal reinforcement at the lower age level while there was no significant difference in the case of the males.

4. Rankings by the teacher of the pupils in terms of the extent to which they are judged to show need for approval show no significant correlations with the response to approval as a reinforcer.

Discussion

Perhaps the most general finding of the study is the extraordinarily poor effect which approval has as a reinforcer compared with physical reinforcement except for girls at the pre-kindergarten level. This finding runs counter to the conception of reinforcement provided in most textbooks on elementary education which generally take the position that adult approval is perhaps the most powerful influence on child behavior. In the case of girls at the pre-kindergarten level this may be true, but it can hardly stand as a generalization. There would seem to be many reasons why approval may well be a much weaker reinforcing contingency than many writers suppose it to be. Some of these need to be given brief consideration.

First, approving remarks are very common events in the life of most children. Teachers can be observed whose stream of behavior in the classroom consists of a flow of comments or gestures of approval. The commonplace nature of such comments may reduce their value as a reinforcer.

Second, since the behavior of pre-school girls is generally more acceptable to the mother than is the behavior of pre-school boys one might well expect that approval would be less effective for young boys than for young girls.

Third, there is a possibility that expressions of approval toward children may be much more satisfying to the adult who gives them than to the child to whom they are directed.

A second matter of interest is the extraordinary effectiveness of the physical reinforcement in this study. This effectiveness could be accounted for in a number of different ways.

First, the delivery of the marble by mechanical means is a novel event. There is an overwhelming amount of evidence that novel events have reinforcing properties on the behavior of a wide range of different creatures. The novelty of the event may, in this particular case, account for its effectiveness as a reinforcer.

Second, in the case of the delivery of a marble, the subject has the experience of acting directly on the environment and actually being the cause of an event. Such a relationship to the environment is considered by many psychologists, including Woodworth (1958), to be one sought after by most higher organisms.

Finally, the point must be made that the highly sophisticated interpretations of the effect of reinforcers on child behavior, such as those which invoke Freudian concepts, would appear to involve constructs which are too remotely related to be of value at this time. A much simpler theory accounting for the nature of events that reinforce child behavior would appear to be needed.

Summary

The relative effectiveness of verbal and physical reinforcements were compared as a function of age and of sex in a knob-pressing task. The sample was composed of twenty males and twenty females at each of five age levels, representing pre-kindergarten, kindergarten and grades three, six and nine. A highly significant difference was found between the reinforcing treatments with physical being more effective than verbal reinforcement at all age levels. These results had been predicted in the two oldest age groups, however, hypotheses made for the pre-kindergarten and third grade groups were not confirmed. It had been predicted that verbal reinforcement would be more effective than physical with the pre-kindergarten group and the two reinforcements would be equally effective with the third grade group. Giving a verbal reinforcement simultaneously with a physical reinforcement resulted in no greater response than did the physical reinforcement alone.

Pre-kindergarten females respond significantly more to verbal reinforcement than did the third grade females, while no difference was demonstrated between male Ss at the two age levels.

References

Alberts, E., and Ehrenfreund, D. Transposition in children as a function of age. J. exp. Psychol., 1951, 41, 30-38.

Brackbill, Y., and Jack, D. Discrimination learning in children as a function of reinforcement value. Child Develpm., 1958, 29, 185-190.

Gewirtz, J. L. Three determinants of attention-seeking in young children. Monogr. Soc. Res. Child Develpm., 1954, 19, No. 2 (Serial No. 59), 5-46.

Jones, M. H., and Liverant, S. Effects of age differences on choice behavior. Child Develpm., 1960, 31, 673-680.

Kuenne, M. R. Experimental investigation of the relation of language to the transposition behavior in young children. J. exp. Psychol., 1946, 36, 471-490.

McCullers, J. C., and Stevenson, H. W. Effects of verbal reinforcement in a probability learning situation. Psychol. Rep., 1960, 7, 439-445.

Meehl, P. E. On the circularity of the law of effect. Psychol. Bull., 1950, 47, 53-75.

Page, E. B. Teacher comments and student performance: A seventy-four classroom experiment in school motivation. J. educ. Psychol., 1958, 49, 173-181.

Ro nblith, Judy F. Learning by imitation in kindergarten children. Child Develpm., 1959, 30, 69-80.

Stevenson, H. W., and Cruse, D. B. Effectiveness of social reinforcement with normal and feeble-minded children. J. Pers., 1961, 29, 124-135.

Terrell, G. The role of incentive in discrimination learning in children. Child Develpm., 1958, 29, 231-236.

Woodworth, R. S. Dynamic Psychology. New York: Henry Holt & Co., 1958.

SECTION III

CHAPTER 2

COMPARISON OF THE REINFORCING EFFECT OF DIFFERENT STIMULI IN A PAIRED ASSOCIATE LEARNING TASK

Although strong aversive stimuli have been used extensively to control the learning of subhuman subjects, relatively little research has been undertaken with the stronger forms of this class of stimuli as negative reinforcers of human behavior.

The well-known study by Warden and Aylesworth (1927) provided results showing that rats learned a visual discrimination task nine times faster when rewarded (food) for correct responses and punished (shock) for incorrect responses than when they were rewarded alone. Although the results of the latter study have important implications both theoretical and practical in character, only within the last few years have there been attempts to compare the effects of similar conditions with human subjects. Brackbill and O'Hara (1958) attempted a similar experiment with kindergarten children to compare the effects of reward (candy) alone with the effect of a combination of punishment (taking away of candy). The results of the Brackbill and O'Hara study were in the same direction as those of Warden and Aylesworth but much less dramatic though statistically significant.

In another study by Penney and Lupton (1961) children of elementary school age received 60 trials on a two-choice discrimination problem. One group received rewards (candy) only. A second group was punished (98db tone) for incorrect responses. A third group was rewarded for correct responses and punished for incorrect responses. In this case the punishment alone condition led to faster learning than the punishment plus reward condition. An odd and unexplained fact about the latter study is that reward produced no observable learning effect while the conditions involving punishment did. Penney and Lupton explained their

results in terms of a drive resulting from the presence of the intense tone and the associated frustration.

The studies which have been cited do not include the use of severe punishment such as may be produced by electric shock, which has been typically used in animal experiments. The effect of shock on the learning of children is quite unknown. The studies also provide no information concerning the relative effects of shock as compared with some of the more conventionally administered signals such as "wrong" or "right." Knowledge concerning the latter is pertinent with respect to the issue of the reinforcing value of events which are strictly informational in contrast with the reinforcing events which also produce strong approach or avoidance behavior. Events which are typically considered as reinforcers of human behavior generally have high information content but do not produce strong approach or avoidance behavior when used out of context. While the word "right" is said to be a powerful reinforcer of behavior involving problem-solving situations, it is not likely to produce approach or avoidance behavior when it is emitted at random by a machine. In contrast, small pieces of candy are likely to produce vigorous approach behavior even when they are not contingent upon particular behaviors. The difference in the operation of these two classes of reinforcing events in the case of human learners has not been explored.

In the two experiments that follow a comparison is made between three reinforcing conditions, the words right or wrong, a tone to indicate either correct or incorrect responses, and an electric shock to indicate either correct or incorrect responses. One hypothesis was that a reinforcing event such as shock, which has a high value for evoking avoidance

behavior, when contingent upon incorrect responses will produce more rapid learning than the word "wrong," which in turn will produce more rapid learning than the tone (which does not ordinarily produce avoidance behavior). In addition it was hypothesized that a reinforcer, such as shock, which produces avoidance behavior will be much more effective in producing learning when it is used as a signal for wrong than when used as a signal for right. It was further hypothesized that information given regarding right responses will be more effectively utilized than information given regarding incorrect responses.

In addition, the Children's Manifest Anxiety Scale developed by Casteneda (1956) was administered to all subjects. It was hypothesized that the high anxiety groups would show less learning than the low anxiety groups and that the high anxiety subjects would show less learning under experimentally induced anxiety conditions than the low anxiety groups.

EXPERIMENT I

Method and Procedure

In developing a procedure for testing the hypotheses, two major assumptions were made: first, a well-established procedure from which most of the technical pitfalls have been eliminated should be utilized and, second, the task should be one which bears a relationship to learning as it commonly occurs in schools. A paired associate learning task satisfies the first of these two criteria and it also simulates in many important characteristics typical school learning and recitation procedures. In representative paired associate learning tasks, the procedure involves the following steps:

1. Both members of the pair are presented and the subject reads them or otherwise responds to them.
2. The first member is presented alone and the subject attempts to reproduce the second member of the pair. Reinforcement may or may not be immediately given.
3. The subject is exposed to a new learning trial with both members of the pair present.

This procedure is similar in many important respects to a typical recitation procedure which proceeds along the following lines:

1. The pupil learns that the capital city of Spain is Madrid.
2. The teacher asks the pupil "What is the capital of Spain?" (The teacher may then provide reinforcement with such words as "right" or "wrong," but she may skip to the third step).
3. The teacher repeats "The capital of Spain is Madrid."

The procedure which has been described permits learning regardless of the presence or absence of the reinforcers. Since learning in this situation is not dependent upon the occurrence of the reinforcing events, it seems desirable to also study the effects of the reinforcers in a situation in which learning was dependent upon the information provided by the reinforcing contingencies. The second experiment, described later, studied the effects of the different reinforcing conditions on a task in which learning could take place only as a result of the reinforcing events.

Design

The study involved three different reinforcing contingencies, an electric shock, verbal feedback, and a tone of approximately 400 cycles

per second. Reinforcement was provided either for right responses or for wrong responses. In addition, the data permitted an analysis of those scoring high on the anxiety scale in contrast to those scoring low. Thus, the variables formed the basis for a 2 X 2 X 3 factorial design.

The subjects were randomly assigned to one of six groups: (1) shock-right, (2) shock-wrong, (3) tone-right, (4) tone-wrong, (5) oral-right, (6) oral-wrong. This placement was accomplished by having each subject draw a token from a small box. The first subject had six conditions to draw from, the second subject only five conditions, etc. The drawing continued, as each subject presented himself, until the tokens representing the six conditions had been depleted. After each group of six subjects, the tokens were returned to the box, and the drawing was begun again. This continued until six groups of thirty subjects had been filled.

Tasks

The tasks used in this study were essentially the same as those developed and first used by Ruch (1934). Ruch used the tasks as part of a large battery of tests comparing the learning ability of the aged and the young. Korchin and Basowitz (1957) used the tasks in a partial replication of the Ruch study. Two of the three tasks were also used by Korchin and Levine (1957) who considered the relationship of anxiety to verbal learning. Data compiled by both Ruch and Korchin gave information on how readily the lists were learned by young and old Ss. Korchin also gave comparative information on the learning of these tasks by normal and psychiatric Ss.

A preliminary run of subjects revealed that Task I was much too easy. The pairs were subsequently scrambled to reduce associative strength.

This was done in order that shock-right and shock-wrong modes of reinforcement would not result in a highly disproportionate number of shocks for shock-right subjects.

As used by Ruch, Task I was:	As scrambled and used in this study Task I was:
1. stem-bud	1. stem-chair
2. house-visit	2. house-bud
3. room-light	3. room-owl
4. soft-chair	4. soft-flag
5. tree-flag	5. tree-car
6. nest-owl	6. nest-light
7. white-pink	7. white-visit
8. walk-car	8. walk-pink
9. horse-sheep	
10. wind-brook	

Only eight of the ten original pair were used, as in the Korchin studies. The three tasks used in the study, including the modification of Task I just shown, are listed in Table 2.01.

The tasks are presumed to differ in difficulty and novelty as well as in the degree to which prior experience influences present learning.

The method selected for the experiment was paired associate learning. The subjects were required to respond to a stimulus word, letter, or number combination by anticipating a paired word, letter, or number. The pairs were presented in changed order, on successive presentations after the second, to prevent serial learning.

Lists were presented using the memory drum operating at a three-second

rate with a three-second blank exposure between the stimulus and response presentation. A short time was allowed between tasks, with a one minute rest between Task II and the final one.

TABLE 2.01

PAIRED ASSOCIATE LEARNING TASKS USED IN THE STUDY

Task I	Task II	Task III
1. stem-chair	1. r x s = q	1. 5 x 5 = 11
2. house-bud	2. h x y = j	2. 3 x 1 = 1
3. room-owl	3. b x d = m	3. 3 x 3 = 4
4. soft-flag	4. a x m = b	4. 3 x 4 = 2
5. tree-car	5. s x q = h	5. 5 x 1 = 7
6. nest-light	6. k x u = t	6. 6 x 3 = 5
7. white-visit	7. w x l = s	7. 2 x 5 = 8
8. walk-pink	8. l x b = d	8. 2 x 4 = 9

No practice list was used. The subject was presented the list the first time through and told he would be required to learn the second word of each pair. Responses were recorded for the second and subsequent trials.

The lists were always presented to the subjects in the same order. While there would be advantages in counterbalancing the order, there is also an advantage in administering the simpler task first, particularly in those treatments involving electric shock. For this reason the order of the task is constant, with the simplest of the three tasks given first.

Apparatus

The learning lists were mechanically presented by a Gerbrand's memory drum, Model M1-A. An elongated paper cylinder, on which were typed six orders of the eight pairs making up each learning task, was taped over each of the three drums and was used with a large faceplate and sliding mask. This arrangement proved superior to the more usual endless paper tape presentation method which is often plagued with problems of synchronization, transport, and alignment. The paper cylinder allowed quick and easy changing of tasks and obviated synchronization, transport, and alignment difficulties.

The device used to present the shock reinforcement is shown schematically in Figure 2.01. This homemade shock source allowed the use of either rapidly pulsing or single-discharge shock stimuli. The former mode was used in this study as the young subjects were less fearful of a "tickle shock" than of a single condenser-discharge shock of the same intensity. The wave form and intensity characteristics of the shock stimuli are shown in Fig 2.02.

The tone used as one reinforcement mode, was provided by a simple 400 c.p.s. audio signal generator, shown schematically in Figure 2.03. This audio generator fed a ten watt audio amplifier, which in turn drove one of a pair of PDR-10 Permaflux headphones. The other phone was replaced with a small pad of gauze to keep the phone mount from irritating the subject's head. The level of the tone was set to provide a loud, but not aversive reinforcement condition.

A third reinforcement mode using verbal reinforcement conditions was provided by the experimenter pronouncing the words "right" or "wrong" as

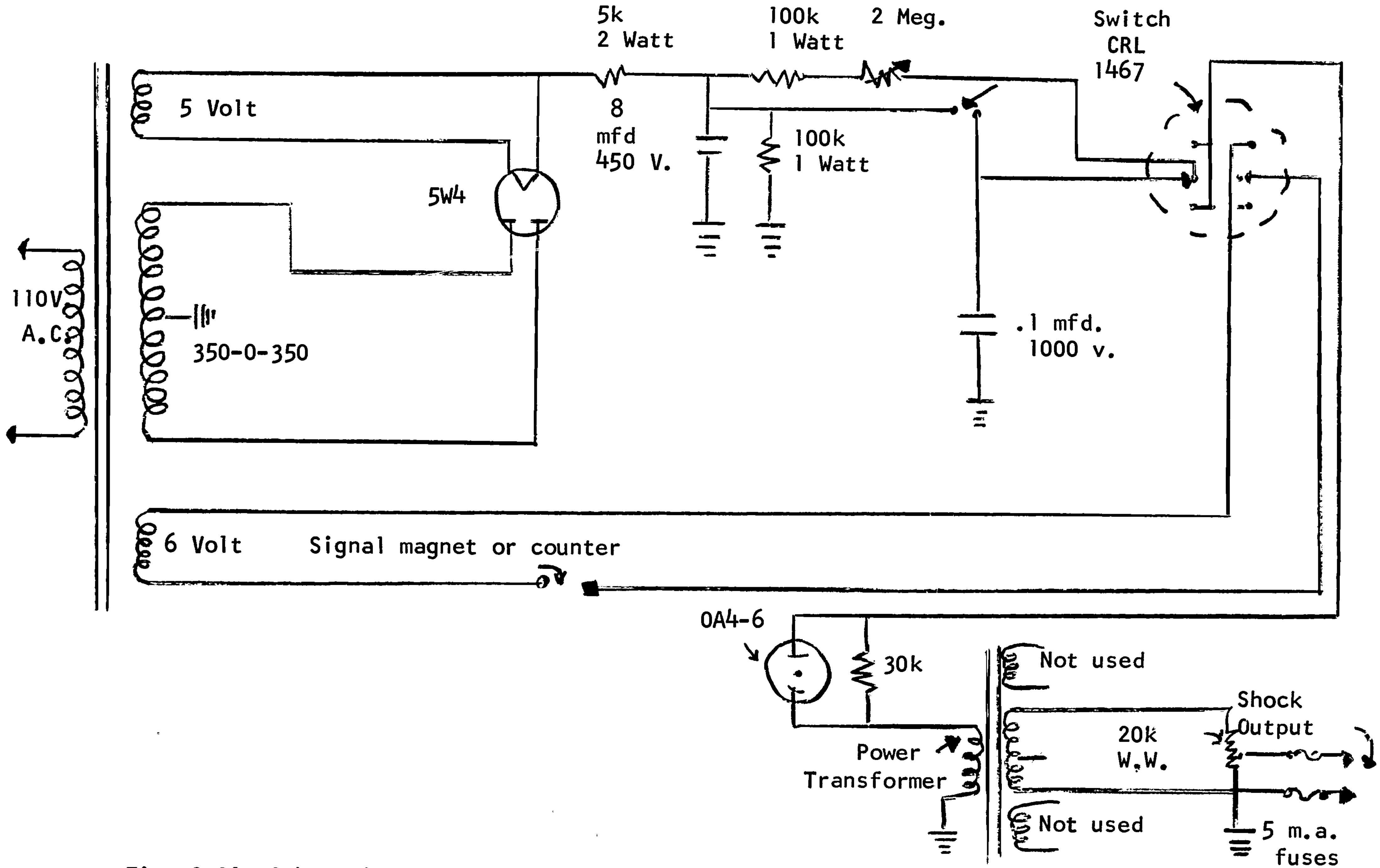


Fig. 2.01 Schematic diagram of shock-device.
Designed by A. B. Cullen. Science, 1941, 94, 219-220.

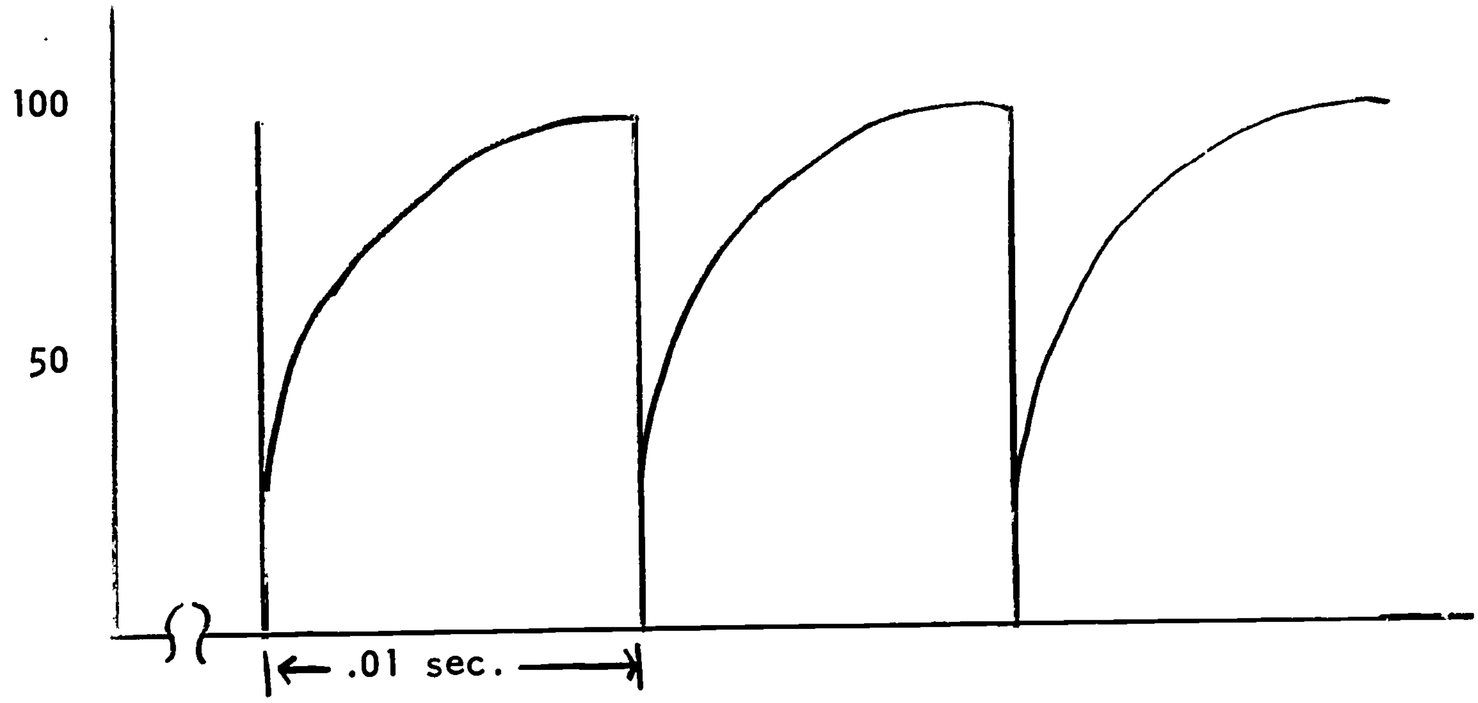


Figure 2.02 Wave form and intensity of electric shock used as reinforcing agent in the study. Instantaneous peak current flow of 4 milliamperes through skin of subject showing resistance of 25 KΩ.

(Techtronix Model 585 oscilloscope used to display the wave form and to determine its temporal and intensity characteristics).

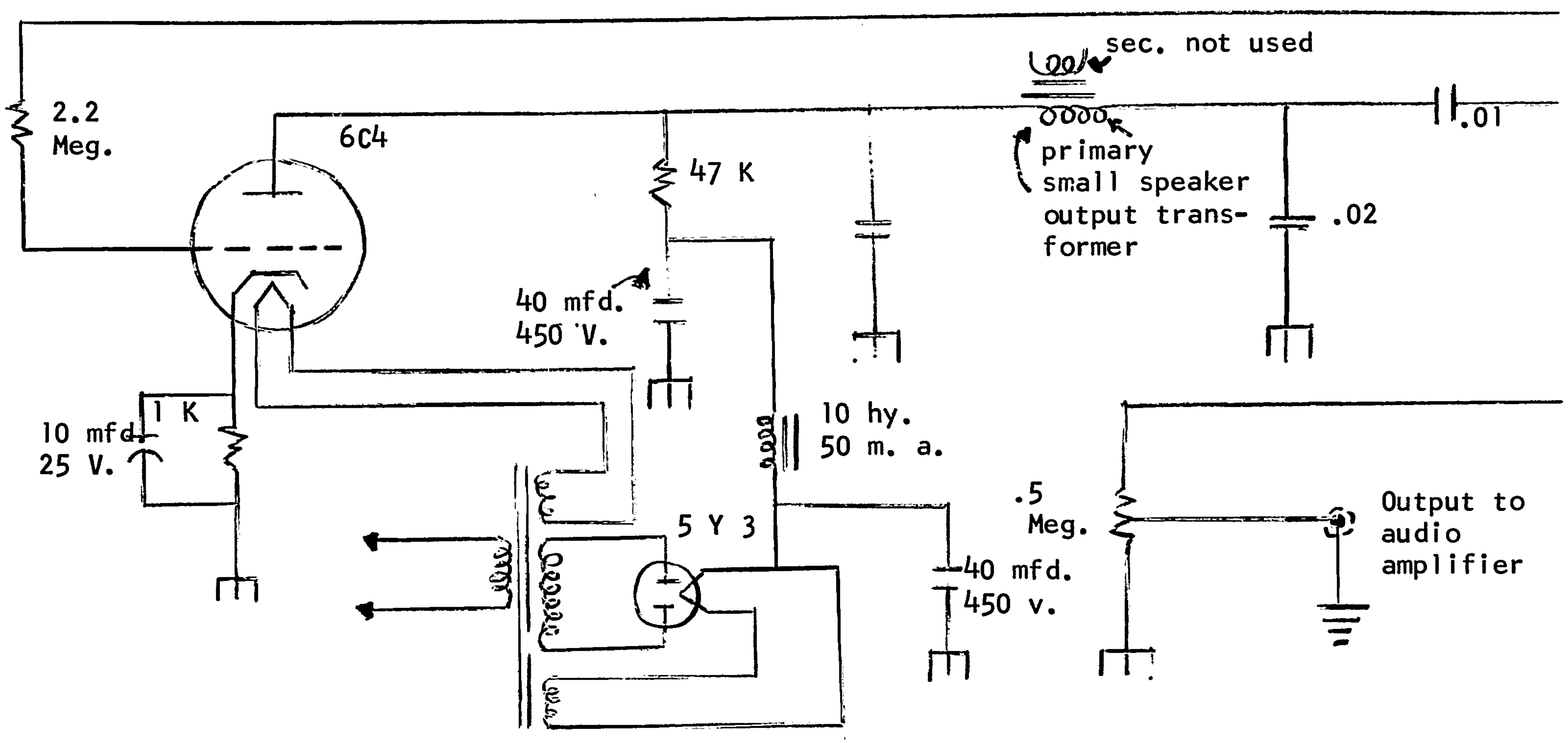


Fig. 2.03 Schematic diagram of fixed-frequency audio oscillator used as tone source. Design from Radio Amateur's Handbook, 1960, 37, 315.

needed to produce the verbal-right or verbal-wrong reinforcement conditions. These reinforcement conditions were uniform for all subjects as a single experimenter handled all the subjects.

Subjects

One hundred-eighty fifth and sixth grade elementary school children served as subjects for the experiment. These subjects were obtained and tested in three locations in Salt Lake City, Utah, during the summer vacation months. Approximately equal number of boys (N=92) and girls (N=88) were tested.

Subjects were selected on the basis of their willingness to participate in the experiment utilizing paired associate verbal materials in which oral, shock, and tone reinforcement modes were used. Each subject was paid $1.00 at the completion of the learning session.

An effort was made to control contamination by asking subjects not to tell others about the learning tasks or procedures. How effectively this precaution prevented contamination can only be surmised.

Procedure During Acquisition

As a subject came into the room where the experiment took place, he was seated in a chair to the experimenter's left. Both S and E faced a long table on which the memory drum, the shock apparatus, the signal generator and audio amplifier, one of a pair of headphones with its head band, a stack of the forms used to keep track of the learning scores and a small box containing wooden tokens were situated. The tokens were used in the assignment of S to a particular treatment. A checklist of the items to be included in the explanations and instructions was on the

table, but was kept for reference only. Instructions and explanations concerning the apparatus and procedure were given in an informal manner, modified somewhat in form with each individual, as some Ss showed more curiosity about one aspect of the experiment, some about another aspect. Enough time was taken with each child so that he had his questions clarified regarding the entire procedure. This was necessary as the ages of the Ss ranged from 10 years three months to 13 years six months. This age range provided a large span as far as sophistication and confidence level was concerned.

The following instructions were given to each subject before Task I:

> "This is one of several experiments being done by people from the University of Utah. We are interested in discovering which are the best ways to help children learn."
>
> "This device, called a memory drum, will automatically show you what is to be learned. There are three different learning tasks. The first one uses common words with which you are very familiar. See this little window? When I turn on the drum, the drum will revolve and a word will appear right here. Read the word aloud so that I can hear it. The word will disappear when the drum moves again. The window will be blank for three seconds and then the word you saw and read aloud will reappear. This time it will have another word with it. Read both words aloud and try to remember them as a pair. They belong together and you are to say the second one whenever the first of the pair appears. There are eight such pairs of words and you will have six trials or chances to learn them."
>
> "On the second time through these eight pairs, when you see the first word, read the word aloud. Then say the first word and the word that belongs with it before the two words come into view on the next turn of the drum."
>
> "Your score will be the total number of times you give the correct response word before it appears with the first word."
>
> "There are six ways you may be shown that you have given the correct or the wrong response. First, you may be given a tone in your ear every time you are right, or you may be given the tone only when you give wrong words or no word at all. A

second way you may be helped to learn these word pairs is to be told the word "right" for each correct word, or to be told "wrong" when you give the wrong word or do not give any reply. The third way you may be helped to learn the word pairs is by being given a "tickle-shock" on the tips of the fingers of your left hand for each right response. Shocks may be given for wrong responses also."

"Let me show you these different ways of helping you learn. This is the tone (placing earphone over the left ear). The tone is loud, but not so loud as to hurt your ear. O. K., let's take the ear phone off and see what the "tickle-shock" feels like. Put your fingers down on these two pieces of metal like this. (Experimenter demonstrates). These are the shock electrodes. The shock is unpleasant. This is how it feels." (Demonstration of shock)

"You will learn using only one of the different ways, not all of them. If you think you would like to take part in this experiment, now that you know what it is all about, draw a token out of this box (experimenter shakes the tokens about in the box, holding it above the S's eye-level) and let us see which way you will be helped to learn the pairs of words."

"You drew S. W. That stands for shock-wrong. Now remember that each time you do not give the correct word, or give no response, you will be given a shock. Let us tape the electrodes on your finger-tips and begin." (Appropriate comments were given for the other five conditions).

"Now we will turn on the drum. Watch the little window and read the word aloud as it comes into view. (E prompts the S as necessary on each of the eight pairs for the first viewing trial). Now, on the second time through the list, you will receive the shock for each wrong word, or for failure to give any word before the paired words appear."

After each trial of the list, the sliding mask was moved over one position, exposing the same eight pairs, but in a changed order.

All responses were recorded on a scoring sheet on which E underlined each correct response and recorded incorrect responses by writing them. Failure to respond at all was indicated by a check.

At the completion of the six trials on task one, the drum's cylinder

was replaced with a second one which was used to present task two. The same procedure, timing, and reinforcement were used.

At the end of Task I the following instructions were given before administering Task II:

"You now know how the memory drum works and how we keep score. This next list is harder than the first one, so a lower score may still be considered a good score. Do the best you can."

"This time you are to learn letters paired like this (E shows S an example of the nonsense equations, E x G = Z). The E x G part will appear in the window first, just as the word did last time. Say aloud, E x G equals, then wait until the drum moves and shows you waht E x G does equal. Then repeat the whole combination aloud. This is called a nonsense equation. You are to learn which letters go together just as you learned which words were mates."

"I will turn the drum on and you read the letters aloud so that I can hear you. Ready? (E prompts S when needed in order to get correct response patterning to occur). Now, the second time through the list, I will keep your score." (Second task is presented).

At the end of Task II, E gave the following instructions:

"That is fine. Let us total up your score for that one. One more list and that completes the learning part of the experiment. This last list is made up of interference equations. For example, you will learn something like 2 x 5 = 25. The answers are wrong, but learn them anyhow. Pronounce the first item that appears in the window aloud. Then, when the entire equation is visible, say both the part in front of and after the equal sign. Remember these pairs belong together as did the words and number combinations."

"Ready? (E prompts the subject to help establish response pattern as in Task I and Task II)."

"That is all of the learning tasks. Let us total your score."

"I would like to know a little more about how you feel in various situations. Before you go, would you mind marking these two pages of statements for me. (The statements referred to here were the items of the Children's Manifest Anxiety

Scale). It may be hard to make up your mind, but in each case the statement will be mostly, or most often true of you, or mostly, or usually false. Mark a T in front of those statements that are mostly, or more often true for you and an F for mostly false. When you have finished both pages, and a few questions at the bottom of the last sheet, you are through. Now, if you will please sign the pay sheet, I will give you your dollar."

"Thank you for helping me. Be sure not to tell other children any of the things you have learned. Let them come and find out for themselves. If you tell someone else an answer, this may damage the usefulness of the experiment and we do not want you to have done all this for nothing."

A number of subjects expressed fear and disinclination to receive the shock reinforcement, but drew a token none the less expressing relief whenever alternative reinforcing modes were drawn. No subject actually refused to participate after he had presented himself at the experimental room. However, as word of mouth advertising was used to secure subjects, extremely fearful subjects may have failed to appear at all. The experimenter asked each subject to provide one additional subject. Most subjects complied with this request.

Results

All scores for performance during acquisition are the number of correct anticipations during six trials. Tables 2.02, 2.03, 2.04 and 2.05 give the means and the standard deviations for the various subgroups for each of the tasks and for the total scores derived from each task. The performance of the various subgroups gives the appearance of high homogeniety from treatment to treatment. The large differences found in studies with animals which compare electric shock with nonaversive reinforcers certainly are not apparent.

Table 2.02

MEAN AND STANDARD DEVIATIONS OF NUMBER OF CORRECT

ANTICIPATION FOR ALL TASKS COMBINED IN EXPERIMENT I. (N=15 FOR EACH CELL)

	High Anxiety				Low Anxiety				Total	
	Right Reinforced Responses		Wrong Reinforced Responses		Right Reinforced Responses		Wrong Reinforced Responses			
	M	SD	M	SD	M	SD	M	SD	M	SD
Tone	74.6	20.30	64.1	17.54	69.9	23.25	62.2	20.42	67.7	20.38
Shock	58.5	18.56	71.1	18.87	71.9	15.88	78.1	19.15	69.9	18.12
Oral	72.1	13.40	71.6	12.27	72.0	11.69	79.4	10.80	73.8	12.04
	High anxiety group M = 68.7				Low anxiety group M = 72.3					
	Right responses reinforced M = 69.9				Wrong responses reinforced M = 71.1					

Table 2.03

MEAN AND STANDARD DEVIATIONS OF NUMBER OF CORRECT ANTICIPATIONS FOR TASKS IN EXPERIMENT I. (N=15 FOR EACH CELL)

	High Anxiety				Low Anxiety				Total	
	Right Reinforced Responses		Wrong Reinforced Responses		Right Reinforced Responses		Wrong Reinforced Responses			
	M	SD	M	SD	M	SD	M	SD	M	SD
Tone	26.9	10.50	26.2	8.85	22.7	11.50	19.9	10.68	23.9	10.38
Shock	21.1	6.22	25.7	6.64	27.7	8.01	28.2	8.55	25.7	7.36
Oral	29.3	9.68	26.5	7.47	27.5	9.91	32.0	7.25	28.8	8.58
	High anxiety group M = 26.0				Low anxiety group M = 26.3					
	Right responses reinforced M = 25.9				Wrong responses reinforced M = 26.4					

Table 2.04

MEANS AND STANDARD DEVIATIONS OF NUMBER OF CORRECT

ANTICIPATIONS FOR ALL TASKS COMBINED (N=15 FOR EACH CELL)

	High Anxiety				Low Anxiety				Total	
	Right Reinforced Responses		Wrong Reinforced Responses		Right Reinforced Responses		Wrong Reinforced Responses			
	M	SD	M	SD	M	SD	M	SD	M	SD
Tone	18.2	7.11	13.5	7.79	20.3	7.83	15.9	8.72	17.0	7.86
Shock	14.1	5.57	19.0	7.16	17.8	9.78	21.1	5.55	18.0	7.02
Oral	17.7	9.05	18.4	6.61	17.6	6.48	18.1	6.70	18.0	7.21
	High anxiety group M = 16.8				Low anxiety group M = 18.5					
	Right responses reinforced M = 17.6				Wrong responses reinforced M = 17.7					

Table 2.05

MEAN AND STANDARD DEVIATIONS OF NUMBER OF CORRECT

ANTICIPATIONS FOR ALL TASKS COMBINED (N=15 FOR EACH CELL)

	High Anxiety				Low Anxiety				Total	
	Right Reinforced Responses		Wrong Reinforced Responses		Right Reinforced Responses		Wrong Reinforced Responses			
	M	SD	M	SD	M	SD	M	SD	M	SD
Tone	29.5	8.44	24.4	7.46	26.9	6.92	26.4	8.08	26.8	7.73
Shock	23.3	7.67	26.4	8.02	26.4	8.53	28.8	8.47	26.2	8.17
Oral	25.1	6.92	26.7	9.39	26.9	7.09	29.3	6.46	27.0	7.47
	High anxiety group M = 25.9				Low anxiety group M = 27.4					
	Right responses reinforced M = 26.4				Wrong responses reinforced M = 27.0					

The distribution of anxiety scores for the 180 Ss who served in this study is shown in Figure 2.04. The distribution of the scores in this study is very similar to that derived from Castaneda's (1956) original standardization data also shown in the same Figure. The mean of our sample appears to be a little higher.

Frequency polygons showing the distribution of CMAS scores for boys and for girls appears in Figure 2.05. The distribution for the sexes indicates a somewhat higher mean value for the girls. Castaneda (1956) also reported a general tendency for the girls to receive higher anxiety scores.

Only one form of the CMAS is available. In order to provide an indication of its internal-consistency, Kuder and Richardson's formula KR21 was applied. This formula provides a coefficient of equivalence not unlike that derived by split-half techniques often used in determining internal-consistency measures when equivalent forms of a test are not available. Cronbach (1960) has discussed the usefulness as well as the limits of this Kuder-Richardson formula. In the present consideration of the CMAS, the derived coefficient of equivalence was found to be .75. This closely approximates reported test-retest reliabilities of the CMAS which Smock (1958) reports to range from .61 to .65, depending on the age and sex of the children to whom it has been given. Smock (1958), citing an unpublished study, reported that with seventh and eighth grade subjects he found slightly higher test-retest reliability than those noted above.

The relationship of anxiety to learning is apparently a complex one. In studies by Matarazzo and Phillips (1955) and Matarzaao et al., (1955), a curvilinear relationship was postulated as existing between performance

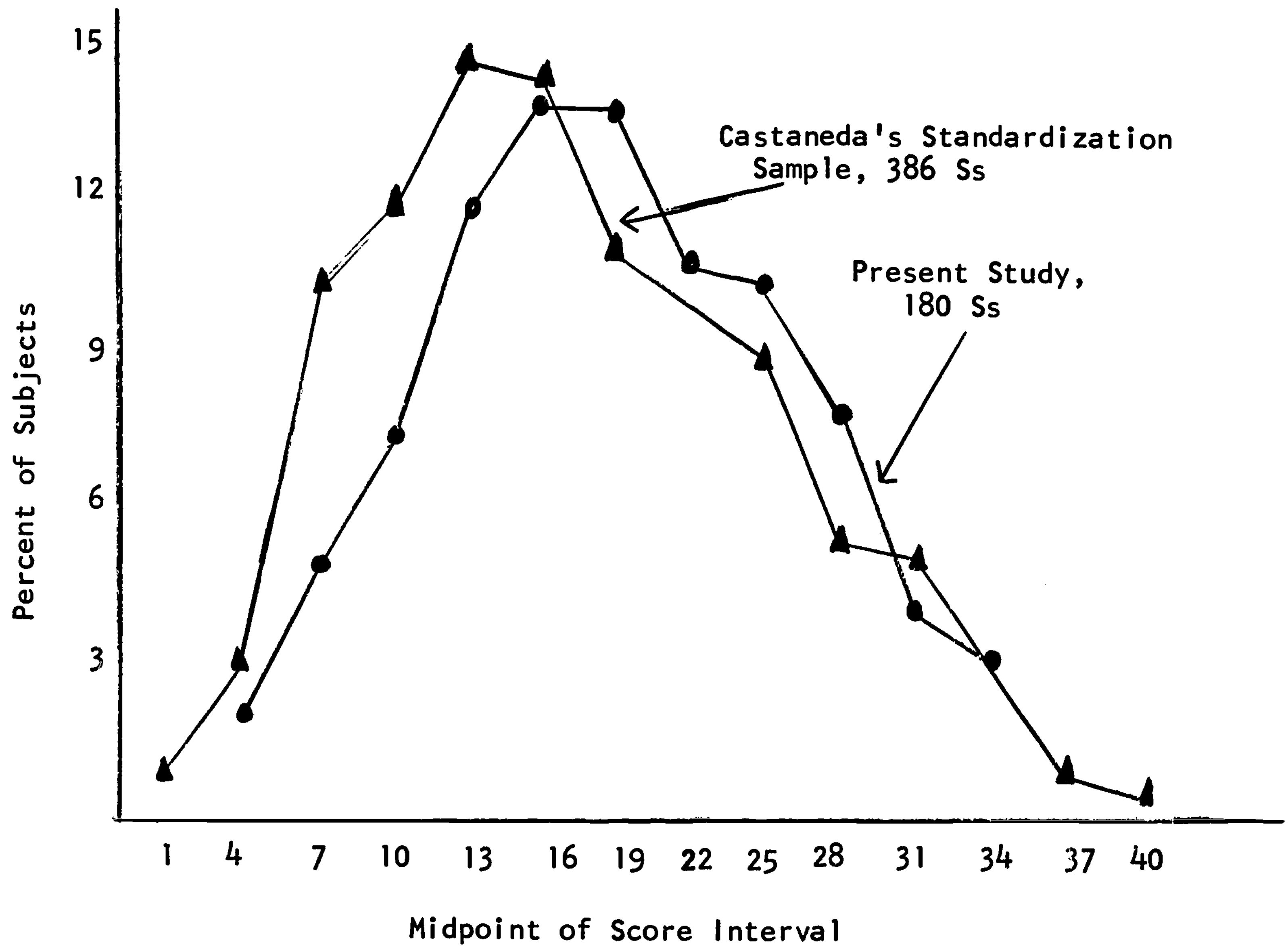


Fig. 2.04 Frequency polygon showing percentage of 180 children receiving the indicated scores on the children's Form of the Manifest Anxiety Scale.

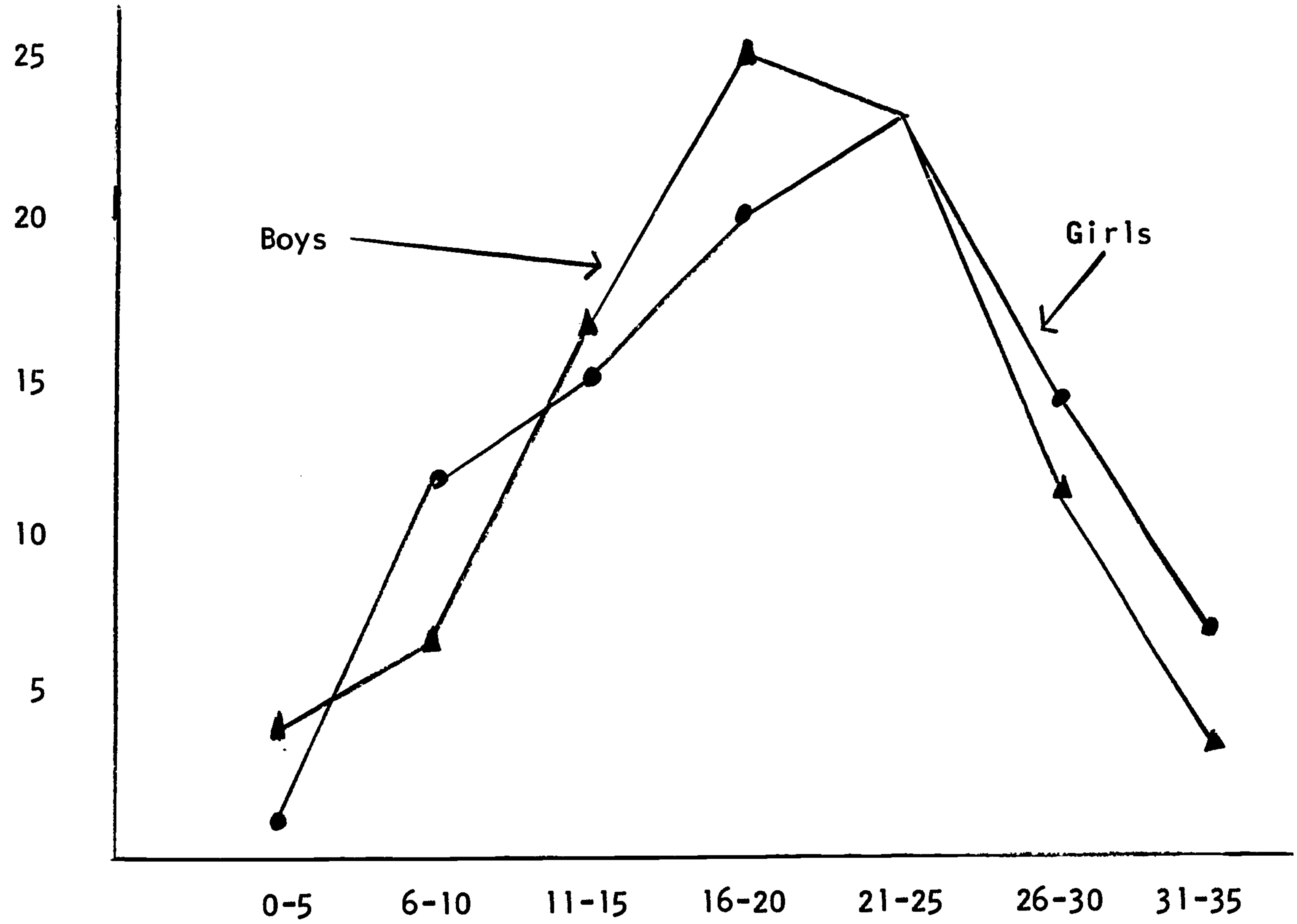


Fig. 2.05 Distribution of Children's Manifest Anxiety Scale score for boys and for girls.

in learning situations and anxiety as measured by a questionnaire. The relationship was demonstrated as significant for timed maze learning but not for digit symbol learning.

A curvilinear relationship might be expected in terms of what is known about drive in relation to learning. As drive is increased, rate of learning also tends to increase; but as drive level exceeds a certain point performance becomes disorganized and declines.

The anxiety scores of the subjects in this study, when plotted on a seven-point scale (shown in Figure 2.06) and compared to learning for each of the three tasks as well as for total learning, failed to show any such curvilinear relationship. This provides a sound basis for the rejection of the hypothesis which postulated that such a curvilinear relationship would be found existing between learning on the tasks and levels of anxiety.

The lack of a consistent relationship between anxiety and learning is emphasized by the correlations shown in Table 2.06. For a correlation to be significant at the .05 level, r must be 0.35 when N=30. Inspection of Table 2.03 shows that only the correlation of CMAS scores and learning with shock reinforcement for right responses on Task I reaches this level. Since one correlation in 20 could be expected to reach this magnitude by chance, one must conclude that the relationship between CMAS and task performance does not reach a significant size.

A number of different analyses of variance may be used to advantage to bring out relevant relationships. What may have been an obscure or insignificant relationship in one analysis may emerge as clear-cut and significant when the variables are analyzed using a different breakdown of the data.

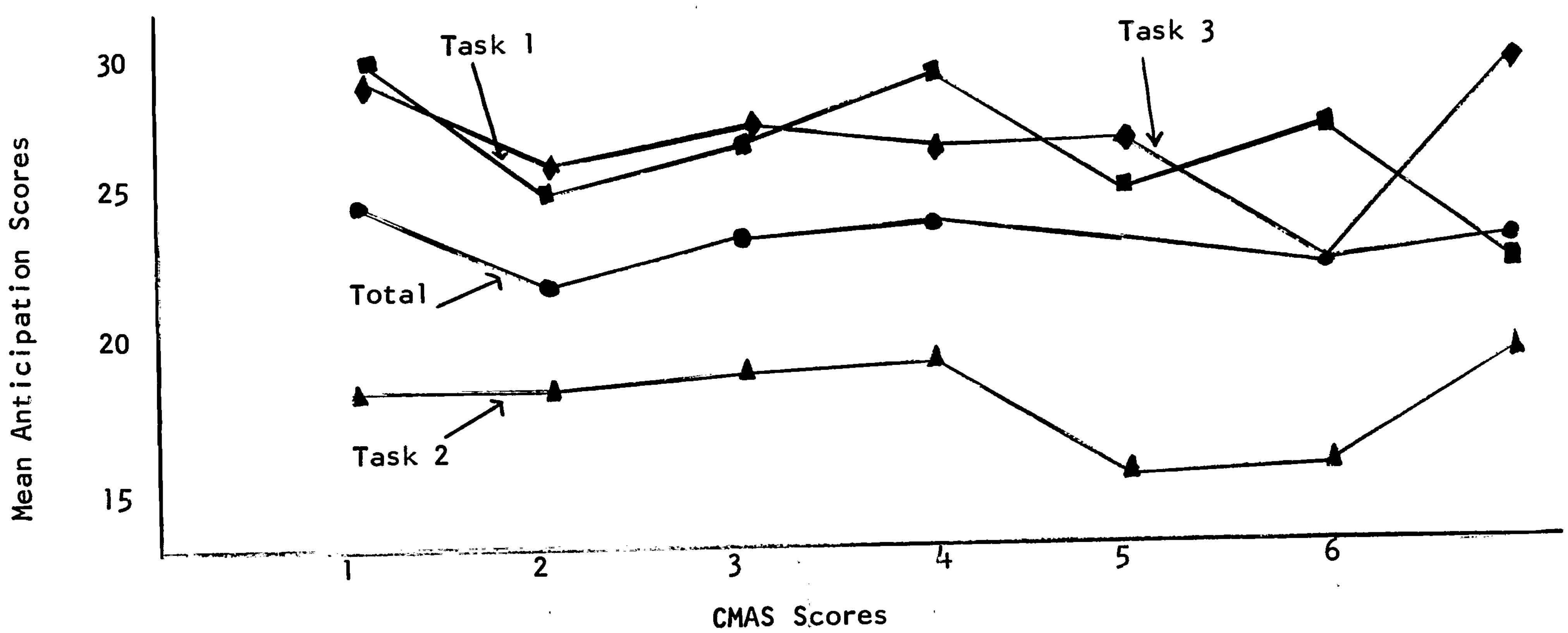


Fig. 2.06 Mean learning scores for each of seven groups ranging from low anxiety (group 1) to high anxiety (group 7).

Table 2.06

CORRELATIONS OF CMAS WITH NUMBER OF CORRECT ANTICIPATIONS IN THE THREE EXPERIMENTAL TASKS. CORRELATIONS ARE REPORTED SEPARATELY FOR THE SIX EXPERIMENTAL CONDITIONS.*

	Task I	Task II	Task III	Task IV
Oral Wrong	-.21	.11	-.06	-.08
Shock Wrong	.00	-.12	-.17	-.11
Tone Right	.01	-.23	-.02	-.09
Oral Right	.12	-.07	-.01	.03
Shock Right	-.35	-.15	-.21	-.29
Tone Wrong	.23	-.14	-.09	.03

*N=30, r required for significance at .05 level = 0.35

A drawing showing the variables and their relationship is presented in Figure 2.07. The three columns represent reinforcement modes---oral, shock, and tone. The two rows include reinforcement of right responses and reinforcement of wrong responses. The layers consist of high and low anxiety levels.

An analysis of variance for these variables using total learning scores is summarized in Table 2.07. The variable involved was the number of correct responses over all learning trials for all tasks. This analysis points to a significant interaction effect of reinforcing of right and wrong responses on the one hand and reinforcement mode on the other, but

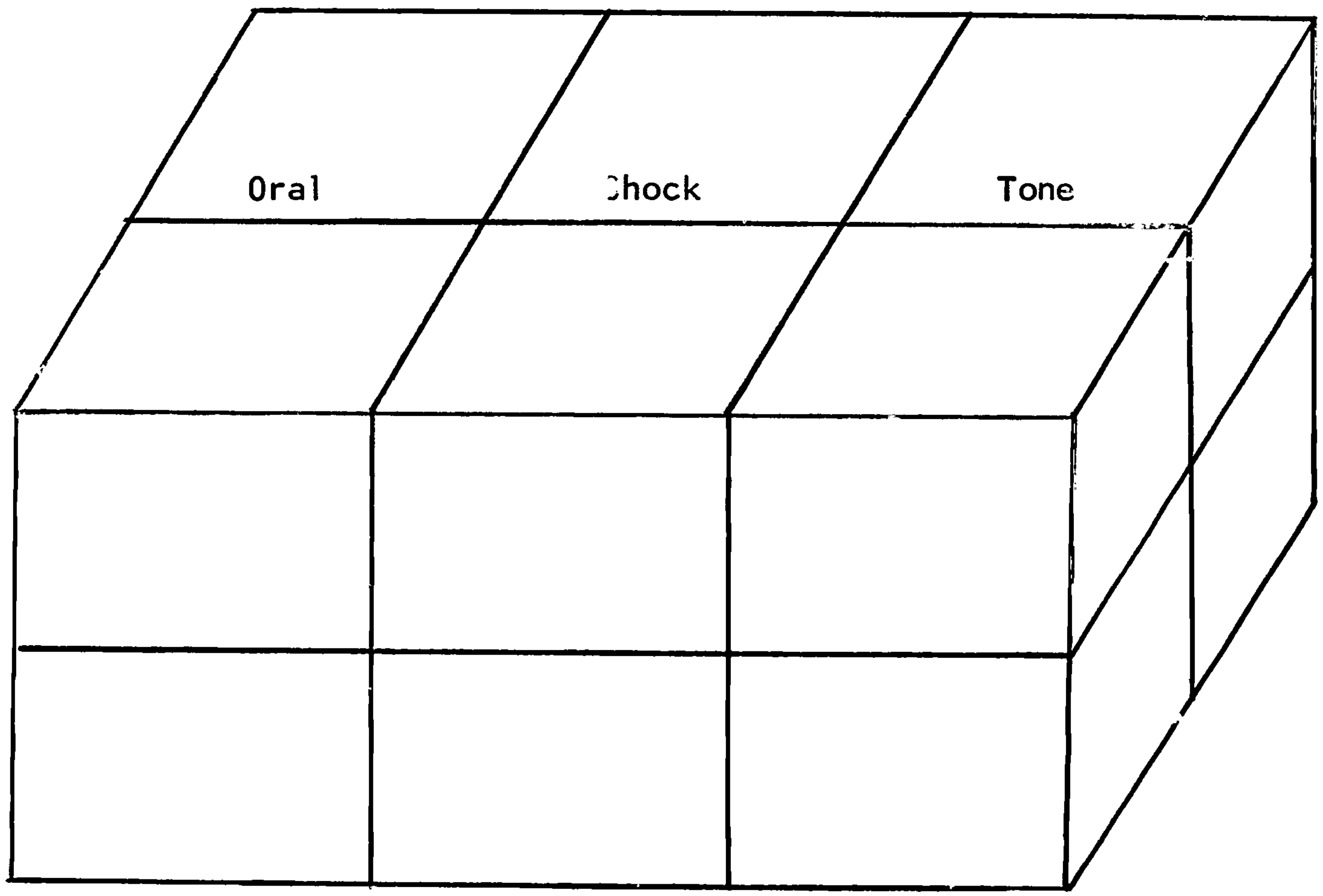


Figure 2.07. Diagramatic representation of treatment combinations for the 2x2x3 factorial design.

TABLE 2.07

ANALYSIS OF VARIANCE FOR TOTAL NUMBER OF CORRECT RESPONSES TO ALL TASKS IN EXPERIMENT I.

Source	Sums of Squares	df	Mean Square	F	p
Right-Wrong	67.22	1	67.22		
Reinforcement Modes	1134.18	2	567.09		
High-Low Anxiety	583.12	1	583.12		
Right-Wrong Reinforcement Modes	2708.04	2	1354.02	3.56	.05
Right-Wrong and High-Low Anxiety	86.27	1	86.27		
Reinforcement Mode and High-Low Anxiety	1362.81	2	680.90		
Right-Wrong Reinforcement Mode and High-Low Anxiety	330.40	2	165.2		
Within Cells	63875.87	168	380.21		
Total	70146.91	179			

reference to Table 2.01 does little to clarify the nature of this interaction and raises the question whether it is anything more than an oddity in this experiment. Anxiety does not appear as a significant main effect in this analysis, and neither does it enter into a significant interaction.

In the analysis as summarized, possible differences due to inter-task difficulty level were obscured by considering only total learning

for the three tasks. In order to further probe for significant interactions and differences, three additional analyses of variance were undertaken using a learning score derived from one task at a time. These analyses are shown in Tables 2.08, 2.09, and 2.10. The same 2 x 2 x 3 factorial design is involved, but learning is considered for each task separately rather than for the scores of all tasks added together.

Table 2.08 shows oral, shock, and tone reinforcement methods to differ significantly in their roles as reinforcing agents. This was to be expected in terms of studies involving subhuman species. However, the same main effect is not found in the analyses of variance for Task II and III, a fact which minimizes any importance which can be attributed to the finding with respect to Task I.

A second effect which can be noted from the analysis shown in Table 2.08 is the significant interaction between anxiety levels and reinforcing modes. Not only do the modes differ significantly but when anxiety and modes interact, the interaction is also significant. This interaction appears to represent the fact that the high anxious group was particularly affected by electric shock and in a negative direction.

Table 2.09 shows that in Task II, learning of non-meaningful equations, a single significant interaction between response modes and reinforcement of right-wrong responses emerges. At this point it is appropriate to recall that Task II proved to be much the hardest of the tasks, the means being: Task I, 26.1; Task II, 17.7; Task III, 26.7. This high level of task difficulty may account for differences in interactions shown to be significant for Task I and not for Task II. Thus, anxiety had no significant effect on learning of Task II.

TABLE 2.08

ANALYSIS OF VARIANCE FOR TASK I IN EXPERIMENT I.

Source	Sums of Squares	df	Mean Square	F	p
Anxiety	1.07	1	1.07		
Right-Wrong	1.96	1	1.96		
Reinforcement Modes	123.47	2	61.74	4.24	.05
Anxiety and Right-Wrong	.30	1	.30		
Anxiety and Reinforcement Modes	128.36	2	64.18	4.40	.05
Right-Wrong and Reinforcement Modes	23.79	2	11.90		
Anxiety, Right-Wrong and Reinforcement Modes	46.16	2	23.08		
Within Cells	2446.93	168	14.57		

TABLE 2.09

ANALYSIS OF VARIANCE FOR TASK II IN EXPERIMENT I.

Source	Sums of Squares	df	Mean Square	F	p
Anxiety	20.83	1	20.83		
Right-Wrong	.01	1	.01		
Reinforcement Modes	6.79	2	3.39		
Anxiety and Right-Wrong	.37	1	.37		
Anxiety and Reinforcement Modes	13.51	2	6.76		
Right-Wrong and Reinforcement Modes	95.05	2	47.52	5.25	.05
Anxiety, Right-Wrong and Reinforcement Modes	1.18	2	.58		
Within Cells	1520.68	168	9.05		

TABLE 2.10

ANALYSIS OF VARIANCE FOR TASK III IN EXPERIMENT I.

Source	Sums of Squares	df	Mean Square	F	p
Anxiety	18.14	1	18.14		
Right-Wrong	3.33	1	3.33		
Reinforcement Modes	3.36	2	1.68		
Anxiety and Right-Wrong	4.54	1	4.54		
Anxiety and Reinforcement Modes	9.42	2	4.71		
Right-Wrong and Reinforcement Modes	13.31	2	6.66		
Anxiety, Right-Wrong and Reinforcement Modes	46.55	2	23.28		
Within Cells	1717.22	10.22			

Turning to the analysis of variance shown in Table 2.10, which provides an analysis of the data for the third task, no significant differences or interactions appear among the variables under observation.

In the previous analysis of variance, scores derived from all trials were summed to produce a total learning score. Some advantage is to be gained by using the separate scores from each trial since this permits the exercising of control over one source of variance which has not been controlled in previous analyses. Three additional analyses of variance were undertaken which involved between trials variance. These are shown for the three tasks in Tables 2.11, 2.12 and 2.13.

Consider first the analysis for the data derived from Task I shown on Table 2.11. The between trials variance is clearly and obviously significant as it would be expected to be. This is hardly surprising for it simply means that learning has taken place. Two interactions, significant at the five percent level should be noted. First, the trials by anxiety by right-wrong reinforcement interaction comes out clearly. Second, there is a significant interaction of trials by anxiety by mode (oral-shock-tone). This is fairly evident in the data which shows that the shock mode of reinforcement is the one in which the high anxiety subjects have a noticeable depression of performance.

TABLE 2.11

ANALYSIS OF VARIANCE FOR TRIALS LEARNING DATA

FOR TASK I IN EXPERIMENT I

Source of Variation	Sums of Squares	df	Mean Square	F	p
Trials	1443.4	5	288.68	248.8	.001
Trials by anxiety	4.51	5	.90		
Trials by right-wrong reinforcement	9.09	5	1.82		
Trials by reinforcing modes	7.24	10	.72		
Trials by anxiety by right-wrong	13.01	5	2.60	2.24	.05
Trials by anxiety by reinforcing modes	22.39	10	2.24	1.93	.05
Trials by right-wrong by reinforcing modes	8.84	10	.88		
Trials by anxiety by right-wrong by reinforcing modes	15.93	10	1.59	1.30	
Residual	973.60	840	1.16	1.16	

Table 2.12 presents an analysis of variance of trials learning data for Task II. This analysis was unproductive of significant interactions, though the main effects of trials was, of course, clearly significant.

The analysis of variance for trials learning in Task III is shown in Table 2.13. Here are found two further interactions which closely approach significance. Both of these involve the anxiety variable. These triple interactions are (1) trials by right-wrong reinforcement and (2) trials by anxiety by reinforcing modes (oral, shock, and tone). The fact that both of these triple interactions involve a trials term indicates that the effect of anxiety is not uniform over the learning procedure.

Lie Scale Scores and Learning

Although anxiety did not emerge as a significant main effect in the study, lie scale scores showed a consistent negative correlation with total learning. Correlations for lie score and total learning for the treatment conditions are shown in Table 2.14. The results agree with those previously reported by Axelrod, Cowen, and Heilizer (1956).

TABLE 2.12

ANALYSIS OF VARIANCE FOR TRIALS LEARNING DATA FOR TASK II IN EXPERIMENT I.

Source of Variation	Sums of Squares	df	Mean Square	F	p
Trials	610.42	5	122.08	95.22	.001
Trials by anxiety	4.70	5	.94		
Trials by right-wrong reinforcement	4.80	5	.96		
Trials by reinforcing modes	9.43	10	.94		
Trials by anxiety by right-wrong	8.36	5	1.67	1.30	(2.22=.05)
Trials by anxiety by reinforcing modes	7.55	10	.76		
Trials by right-wrong by reinforcing modes	12.33	10	1.23	.96	
Trials by anxiety by right-wrong by reinforcing modes	11.22	10	1.12	.85	
Residual	1077.17	840	1.28		

TABLE 2.13

ANALYSIS OF VARIANCE FOR TRIALS LEARNING DATA FOR TASK I IN EXPERIMENT I.

Source of Variation	Sums of Squares	df	Mean Square	F	p
Trials	806.76	5	161.35	131.17	.001
Trials by anxiety	7.70	5	1.54	1.25	
Trials by right-wrong reinforcement	7.10	5	1.42	1.15	
Trials by reinforcing modes	21.99	10	2.20	1.78	(1.84=.05)
Trials by anxiety by right-wrong	6.98	5	1.40	1.13	
Trials by anxiety by reinforcing modes	8.66	10	.87		
Trials by right-wrong by reinforcing modes	12.85	10	1.29	1.04	
Trials by anxiety by right-wrong by reinforcing modes	21.88	10	2.19	1.78	(1.84=05)
Residual	1035.03	840	1.23		

TABLE 2.14

RELATIONSHIP OF LIE SCORE ON THE ANXIETY SCALE TO TOTAL PERFORMANCE ON THE THREE TASKS (N=30, p=0.05 for r=0.35)

Reinforcing Condition	Correlation
Oral wrong	-0.45
Shock wrong	-0.31
Tone wrong	-0.02
Oral right	-0.31
Shock right	-0.22
Tone right	-0.43

EXPERIMENT 2

Method and Procedure

In the previous experiment, subjects could learn each task without the reinforcement following each trial. In the second experiment the tasks were so modified that the reinforcing events were necessary for learning. In order for this to be the case, the response to each stimulus term was modified so that it came from a restricted number of categories. In the case of Task I the correct response to the stimulus word was either the word "long" or the word "short." On Task II the correct response was either A or B and on Task III either 1 or 2. Subjects were told what the alternatives were before each task and at no time was the stimulus presented together with the correct response on the memory drum.

Design

The design was essentially the same as in the previous experiment with either right responses or wrong responses reinforced, with three modes of reinforcement (shock, oral, tone), two levels of anxiety, and also two levels of difficulty of the tasks.

Task

The task involved the learning of a correct response to a series of stimuli. The directions were essentially the same as those used in experiment 1 except that the subject was told in the first task that the answer was always either the word "long" or the word "short." In the case of the second task the response to each stimulus was either "a" or "b" with only one being correct. In the case of the third task the correct

response to each stimulus was either 1 or 2. After each stimulus, the subject guessed which one of the responses was correct and received feedback in accordance with the treatment to which the particular subject had been assigned. In contrast with the previous task the subjects in experiment 2 could learn only through the feedback provided since a correct and complete presentation of the stimulus and response did <u>not</u> follow each presentation. All subjects completed all three tasks in the same order. Each subject was run for six trials on each task and his score was the number of correct responses over all six trials.

Tasks

The tasks used in experiment 2 are shown in Table 2.15. Each task was used in a long and a short form.

Subjects

The subjects were 192 boys and girls with an age range closely similar to those in the previous study. They also were obtained in a residential neighborhood and the experiment was conducted in the basement of one of the houses. Subjects were assigned to experimental treatments by drawing a tag from a hat. The tag indicated the treatment to which they were assigned.

Results

Since in this second experiment, learning cannot occur without the signals or reinforcers, one might have expected substantial differences between the main effects. Tables 2.16, 2.17, 2.18 and 2.19 show the means and standard deviation for the groups representing the various treatments. The tables are presented by task and also for total performance on the three tasks.

TABLE 2.15

TASKS USED IN EXPERIMENT 2.

Task I			Task II		Task III	
Stimulus		Response	Stimulus	Response	Stimulus	Response
stem	-	long	r x s =	a	5 x 5 =	1
house	-	short	b x y =	b	*3 x 1 =	2
room	-	long	b x d =	b	3 x 3 =	2
*soft	-	long	a x m =	b	*3 x 4 =	1
tree	-	short	s x q =	a	5 x 1 =	1
nest	-	long	*k x u =	a	6 x 3 =	2
*white	-	short	*w x 1 =	b	2 x 5 =	1
walk	-	short	1 x b =	a	2 x 4 =	2

*These items were omitted to produce the shorter task.

The means do not show any particularly striking differences between reinforcing modes such as has appeared in data involving rats. Indeed, the differences are very small and inconsistent from task to task. The table of means do suggest that there may be differences between the high and low anxiety groups. While the differences between the means are small, they are consistent from task to task.

In addition, there appears to be a small but consistent difference between the learning of right and the learning of wrong responses with

Table 2.16

MEANS AND STANDARD DEVIATIONS OF NUMBER OF CORRECT RESPONSES FOR ALL TASKS COMBINED

		High Anxiety				Low Anxiety				M for Main Reinforcing Conditions
		Right Responses Reinforced		Wrong Responses Reinforced		Right Responses Reinforced		Wrong Responses Reinforced		
		M	SD	M	SD	M	SD	M	SD	M
Tone	Long Task	164.9	17.72	183.1	16.85	190.0	30.82	185.4	19.80	160.7
	Short Task	132.5	18.53	144.3	12.03	145.0	18.62	140.4	18.69	
Shock	Long Task	170.5	19.30	181.1	21.06	178.4	24.88	184.1	13.36	156.5
	Short Task	135.9	26.77	131.3	16.75	132.5	21.19	138.5	15.72	
Oral	Long Task	177.4	31.67	178.1	27.05	161.6	17.63	182.3	25.14	159.0
	Short Task	130.1	27.55	151.2	18.17	147.5	20.02	143.6	24.10	

High anxiety group M = 156.7	Low anxiety group M = 160.8
Right responses reinforced M = 155.5	Wrong responses reinforced M = 161.9

Table 2.17

MEANS AND STANDARD DEVIATIONS OF NUMBER OF CORRECT

RESPONSES ON TASK I

		High Anxiety				Low Anxiety				M for Main Reinforcing Conditions
		Right Responses Reinforced		Wrong Responses Reinforced		Right Responses Reinforced		Wrong Responses Reinforced		
		M	SD	M	SD	M	SD	M	SD	M
Tone	Long Task	52.5	9.58	61.3	7.59	62.1	14.36	59.6	12.91	51.3
	Short Task	37.4	11.84	46.6	7.27	45.7	11.95	44.9	7.36	
Shock	Long Task	53.1	6.24	56.6	7.44	55.1	9.57	61.6	16.04	49.9
	Short Task	41.8	11.84	43.5	3.78	42.3	8.15	44.9	7.55	
Oral	Long Task	52.6	14.65	60.5	8.28	54.0	9.17	64.4	5.93	51.9
	Short Task	39.5	13.35	50.1	8.39	49.8	10.35	44.4	10.94	

High anxiety M = 49.6	Low anxiety M = 52.4
Right responses reinforced M = 48.8	Wrong responses reinforced M = 53.2

Table 2.18

MEANS AND STANDARD DEVIATIONS FOR NUMBER OF

CORRECT RESPONSES ON TASK II

		High Anxiety				Low Anxiety				M for Main Reinforcing Conditions
		Right Responses Reinforced		Wrong Responses Reinforced		Right Responses Reinforced		Wrong Responses Reinforced		
		M	SD	M	SD	M	SD	M	SD	M
Tone	Long Task	57.3	5.34	59.9	9.88	62.1	10.72	64.1	6.56	52.8
	Short Task	44.8	8.22	44.8	7.25	46.0	5.86	43.4	6.91	
Shock	Long Task	58.3	4.10	61.6	11.04	58.6	8.37	59.6	6.78	50.9
	Short Task	42.8	5.92	42.1	5.06	42.1	7.14	42.0	7.27	
Oral	Long Task	66.0	10.73	56.1	11.72	54.5	9.21	61.9	11.75	52.6
	Short Task	40.6	9.50	47.3	8.43	45.1	6.85	49.1	9.91	

High anxiety M = 51.8	Low anxiety M = 52.4
Right responses reinforced M = 51.5	Wrong responses reinforced M = 52.7

Table 2.19

MEANS AND STANDARD DEVIATIONS FOR NUMBER OF

CORRECT RESPONSES ON TASK III

		High Anxiety				Low Anxiety				M for Main Reinforcing Conditions
		Right Responses Reinforced		Wrong Responses Reinforced		Right Responses Reinforced		Wrong Responses Reinforced		
		M	SD	M	SD	M	SD	M	SD	M
Tone	Long Task	55.1	11.41	62.00	7.48	65.8	13.14	61.6	11.92	56.7
	Short Task	50.4	12.83	52.9	7.32	53.3	7.80	52.4	9.36	
Shock	Long Task	59.1	11.51	62.9	12.10	64.6	12.83	62.9	8.58	55.8
	Short Task	51.4	12.98	45.6	8.81	48.1	10.75	51.6	11.60	
Oral	Long Task	58.8	10.95	61.5	14.02	53.1	10.67	56.0	13.69	54.5
	Short Task	50.0	10.90	53.9	9.76	52.6	9.04	50.1	8.85	

High anxiety M = 55.3	Low anxiety M = 56.0
Right responses reinforced M = 55.2	Wrong responses reinforced M = 56.1

the reinforcement of wrong responses producing consistently slightly better learning.

While the table of means and standard deviations is suggestive of the effects of differences in treatments, the testing of the hypotheses implicit in the design require the use of analysis of variance. The analyses of variance corresponding to the table of means and standard deviations are shown in Table 2.20, 2.21, 2.22, and 2.23. These tables present separately the analyses of variance for each task and for the scores derived from all three tasks combined.

The F tests do not provide a very impressive array of significant differences. The most marked is that related to whether right or wrong responses are reinforced. While the direction of this difference is the same across all three tasks, the effect is seen mainly with Task I. From the data one cannot tell whether this is a result of the nature of the task or whether it is due to the fact that the particular task was given first and that during this task the subject underwent adaptation to the learning situation with a resulting change in response to subsequent tasks.

The F test related to the long-short task comparisons has not been computed since it is of trivial interest. The long task - short task comparison is only of interest in the interactions where it did not produce any significant F.

Discussion

When the results of the study as a whole are considered, one cannot help being impressed with the contrast between the data provided by children in the shock situation with the data derived from studies with rats. While rats learn at a vastly faster rate when shocked for incorrect

TABLE 2.20

ANALYSIS OF VARIANCE FOR ALL TASKS COMBINED FOR

NUMBER OF CORRECT RESPONSES IN EXPERIMENT 2

Source of Variation	Sums of Squares	df	Mean Square	F	p
A (treatment)	558.78	2	279.39		
B (anxiety)	796.26	1	796.26		
C (long-short)	71804.01	1	71804.01		
D (right-wrong)	1982.76	1	1982.76	4.14	<.05
AB	720.82	2	360.41		
AC	1270.82	2	635.41		
AD	254.76	2	127.38		
BC	6.37	1	6.37		
BD	497.29	1	497.29		
CD	218.87	1	218.87		
ABC	851.08	2	425.54		
ABD	381.78	2	190.89		
ACD	65.39	2	32.70		
BCD	174.44	1	174.44		
ABCD	2121.83	2	1060.92		
Within	80439.38	168	478.81		
Total	162665.46				

TABLE 2.21

ANALYSIS OF VARIANCE FOR TASK I FOR NUMBER OF CORRECT RESPONSES IN EXPERIMENT 2.

Source of Variation	Sums of Squares	df	Mean Square	F	p
A (treatment)	140.32	2	70.2		
B (anxiety)	368.52	1	368.52	3.35	<.10
C (long-short)	8829.19	1	8829.19		
D (right-wrong)	918.75	1	918.75	8.35	<.001
AB	19.20	2	9.60		
AC	86.16	2	43.08		
AD	54.03	2	27.02		
BC	17.52	1	17.52		
BD	320.33	1	320.33	2.91	.05
CD	90.75	1	90.75		
ABC	12.20	2	6.10		
ABD	333.82	2	166.91		
ACD	114.40	2	57.20		
BCD	126.75	1	126.75		
ABCD	220.28	2	110.14		
Within	18477.75	168	110.00		
Total	30129.98				

TABLE 2.22

ANALYSIS OF VARIANCE FOR TASK II FOR NUMBER OF CORRECT RESPONSES IN EXPERIMENT 2

Source of Variation	Sums of Squares	df	Mean Square	F	p
A (treatment)	137.89	2	68.95		
B (anxiety)	17.52	1	17.52		
C (long-short)	12033.33	1	12033.33		
D (right-wrong)	63.02	1	63.02		
AB	69.51	2	34.76		
AC	83.32	2	41.66		
AD	20.14	2	10.07		
BC	4.69	1	4.69		
BD	30.08	1	30.08		
CD	.19	1	.19		
ABC	228.66	2	114.33		
ABD	197.89	2	98.95		
ACD	250.91	1	250.91	3.57	<.10
BCD	120.33	2	60.17		
ABCD	286.95	2	143.48		
Within	11810.25	168	70.30		
Total	25354.67				

responses, the children do not. While one might have expected that in the shock-wrong condition the children would have learned more rapidly in order to escape from the shock--and this would appear to be a ready means of escape--the fact was that this group did not perform better than the groups which received no shock. In addition, the group that received shock for correct answers was not particularly slowed down by this procedure despite the fact that the better they learned the more times they received shock. The latter group was aware of the fact that the task was to be continued through six trials even if they gave a perfect performance on earlier trials.

A second point to note is that any significant differences which did appear tended to appear on the first of the three tasks. Since the tasks were not administered in a counterbalanced order one cannot tell whether the effect is a result of the fact that the task yielding significant differences appeared first or whether it was the simplest task. Indirect evidence suggests that the effect was probably due to the fact that the task appeared first and that the significant differences were a result of the difficulties which the subjects had in adjusting to the conditions of learning. Certainly, even those who have worked with rats report that the animals become adapted to electric shock. In this particular study, E had the impression that the children showed a similar tendency to adapt. A general conclusion indicated by the data is that the reinforcing value of an event for the subjects considered here is the amount of information which it contains.

In the present study, the reinforcers contingent upon incorrect responses provided the same amount of information as the reinforcers

contingent upon correct responses. In the second experiment a significant difference was found between the rate of acquisition when wrong responses were reinforced in contrast to the reinforcement of right responses. The direction of this trend is opposite to that found in some studies of concept learning, as for example that of Bruner et al. (1956), which have reported that subjects derived more information from the positive reinforcement of correct responses than from the negative reinforcement of incorrect responses. If a similar finding were to occur in other studies of acquisition it would raise an important question for those designing programmed materials who have commonly taken the position that the program should be arranged so that the subject makes only correct responses. Human subjects appear to be much more capable of learning from their errors than some psychologists have supposed them to be.

References

Brackbill, Y., and O'Hara, J. The relative effectiveness of reward and punishment for discrimination learning in children. J. comp. physiol. Psychol., 1958, 51, 747-751.

Castaneda, A., Palermo, D. S., and McCandless, B. R. Complex learning and performance as a function of anxiety in children and task difficulty. Child Develpm., 1956, 27, 327-332.

Korchin, S. J., and Basowitz, H. Age differences in verbal learning. J. abnorm. soc. Psychol., 1957, 54, 63-68.

Korchin, S. J., and Levine, S. Anxiety and verbal learning. J. abnorm. soc. Psychol., 1957, 54, 234-240.

Matarazzo, J. D., and Phillips, J. S. Digit symbol performance as a function of increasing levels of anxiety. J. consult. Psychol., 1955, 19, 131-134.

Penney, R. K., and Lupton, A. A. Children's discrimination learning as a function of reward and punishment. J. comp. physiol. Psychol., 1961, 54, 449-451.

Ruch, F. L. Effects of age on human learning. J gen. Psychol., 1934, 11, 261-285.

Warden, C. J., and Aylesworth, M. The relative value of reward and punishment in the frustration of a visual discrimination habit in the white rat. J. comp. Psychol., 1927, 7, 117-127.

SECTION III

A FINAL WORD

Although the studies considered in this section of the report provide a wide range of reinforcing events with greatly varying characteristics, the results do not indicate a strong relationship between the nature of the reinforcing events and rate of acquisition of the particular task involved. This conclusion holds even when there were drammatic differences in the reinforcing events as was the case when electric shock was compared with verbal approval indicated by the word "Right."

The reinforcers used in the studies served two functions. First, they provided information concerning a correct or appropriate answer. The different reinforcing contingencies provided equal amounts of information within any one experiment and, hence, comparisons could be made in terms of the other characteristics of the reinforcing event. The second characteristic of the reinforcing event which is commonly considered to have special significance is the capacity it has of producing affect. Although the informational property of a reinforcing event can be readily assessed, its capacity to produce affective responses is not readily determined. The evidence that a reinforcing event has strong affective properties is generally qualitative and verbal and is reflected by such statements as "I don't like that," "Do you have to do that again?" or "That's fun." Other evidence that events have affective properties are shown in the tendency to avoid further events in the same category or to seek out repetition of the events. One could also conceive of obtaining evidence of affective response through records of visceral phenomenon such as blood pressure, sweating, changes in peristalsis and so forth, but such evidences are difficult to collect. In the case of the studies that have been reported, it seems clear that the electric

shock used produces strong negative affective response while the word "Right" might be expected to produce mild positive affect. Though these effects cannot be considered as in equal and opposite directions they can be presumed to be sufficiently disparate for our purposes.

The only striking differences between the reinforcing properties of events are found in the first of the two studies reported in which it was shown that a physical event, involving the delivery of a marble, had greater reinforcing value than approval when all age groups involved were combined. A second finding was that for the very young children the girls showed a more marked tendency to respond to approval than girls in the older age groups. On the other hand, in the study involving electric shock in which the aversive property of this reinforcing event was beyond question there were no notable differences in rate of acquisition in treatments involving this reinforcing event than in treatments involving events which might be considered to produce positive affect. These results are in marked contrast with those derived from animal experiments in which the introduction of shock has been shown to have dramatic effects on the rate of acquisition both in motor learning and perceptual discrimination tasks. The data presented here implies that affective responses associated with reinforcers have vastly greater influence on the learning process in subhuman than in human subjects. In the learning situations considered here, the human responds to the information supplied rather than to other characteristics of the reinforcing event. One suspects that the human would learn more effectively from reinforcers carrying a large amount of information but producing negative affect than from reinforcers providing much less information but producing positive affect.

The implication of these studies for education is that human learning should generally be planned in such a way that the reinforcing events supply information in the clearest and most unambiguous form. Information hidden in a cloak of words of praise or blame may loose part of its impact. Indeed, part of the information may be lost when it is thus encapsulated.