JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 1985, 43, 301-313 NUMBER 3 (MAY)

VERBAL RELATIONS WITHIN INSTRUCTION: ARE THERE SUBCLASSES OF THE INTRAVERBAL?

PHILIP N. CHASE, KENT R. JOHNSON, AND BETH SULZER-AZAROFF

WEST VIRGINIA UNIVERSITY, MORNINGSIDE LEARNING CENTER, SEATTLE, AND UNIVERSITY OF MASSACHUSETTS-AMHERST

Six college undergraduates received programmed concept training on three kinds of intraverbal relations. These relations involved definition, exemplification, and example identification questions. The experimenter presented the questions, the subject answered them in writing, and the experimenter provided specific corrective consequences. After completing the training on a concept, the subject immediately received a test on the concept. The test included novel questions similar to the kind used in training (extension tasks) and question types that were not used in training but which were also considered intraverbal relations (transfer tasks). Training results indicated rapid, errorful responding on example identification tasks and slow, accurate responding on exemplification and definition tasks. Test results indicated rapid, errorful responding on example identification extension tasks; slow, accurate responding on exemplification extension tasks; and slow, errorful responding on definition extension tasks. In testing, differential responding occurred on transfer tasks as a function of the kind of intraverbal training received, and substantially lower levels of performance were obtained on transfer tasks than on extension tasks. It appears that the intraverbal can be subdivided into more specific categories of operants.

Key words: verbal behavior, programmed instruction, transfer of learning, written answers, undergraduates

The operant model of language acquisition and change defined by Skinner (1957) has generated considerable discussion over the past 28 years (e.g., Chomsky, 1959; MacCorquodale, 1970; Place, 1981a, 1981b, 1982). Whereas Skinner (personal communication, October, 1980) has called *Verbal Behavior* his most important work, others have argued that the text laid bare the fatal flaws of the behavioral paradigm (Taylor, 1972). While the debate continues, it seems critical to pursue the question: What do the data indicate? Experimenters have investigated the sufficiency of reinforcement for verbal learning (e.g., Azrin, Holz, Ulrich, & Goldiamond, 1961; Frisch & Schumaker, 1974; Greenspoon, 1955; Guess, Sailor, Rutherford, & Baer, 1968; Hart & Risley, 1968; Lee, 1978; Sailor, 1971; Wheeler & Sulzer, 1970), the difference between rule-governed and contingency-shaped behavior (e.g., Baron, Kaufman, & Stauber, 1969; Catania, Matthews, & Shimoff, 1982; Galizio, 1979; Harzem, Lowe, & Bagshaw, 1978), and some specific predictions implied by Skinner (e.g., Lee, 1981a).

The reviews that integrate some of these areas of research (Holz & Azrin, 1966; Lee, 1981b; Segal, 1977) indicate that the operant model does generate testable experimental questions about verbal behavior, and that certain components of the operant model are important to any description of language. These reviews also show that many questions remain to be answered. The present paper addresses one such question regarding whether and how the "intraverbal" (Skinner, 1957) may be divided into more specific operants.

The description of functional classes is a

This paper is based on a dissertation submitted to the University of Massachusetts in partial fulfillment of the requirements for a Ph.D. The authors wish to thank Raymond Lyko and Sandra Gordon for their assistance in conducting this experiment and calculating the data, and the students of Introductory Psychology 190 for their patient participation as subjects in this experiment. Send reprint requests to Philip N. Chase, Department of Psychology, West Virginia University, P.O. Box 6040, Morgantown, West Virginia 26506-6040.

significant feature of Skinner's analysis which goes beyond specific instance-by-instance associations among stimuli and verbal behavior. Other behavioral accounts have emphasized the physical similarities among stimuli and among responses to account for learned relations (see Osgood, 1949). Skinner proposed that the environmental effect of a response defines the relations among stimuli and verbal behavior. If, in correlation with a given set of stimuli, repeated instances of a response have similar effects, then the relations among stimuli, responses, and consequences can be categorized as an operant or a functional class.

This analysis suggests at least one fruitful kind of investigation: Manipulate different interactions among stimuli and responses and categorize those interactions according to their similar effects. We have located only two such studies. Boe and Winokur (1978) manipulated the experimenter's use of specific terminology and found that subjects used particular terms at higher frequencies when the experimenter had used them previously. Skinner (1957) defined this relation as echoic. Lamarre and Holland (1985) manipulated the subjects' requests and identifications of physical objects. Skinner (1957) defined these relations as mands and tacts, respectively. Such experimental analyses help to determine whether Skinner's classification system is useful for organizing the study of verbal behavior, and whether the relations proposed by Skinner are, in fact, functional. Similar studies could investigate other general verbal operants, like the tact, or more specific verbal operants—for example, whether the intraverbal relation can be divided into functional subclasses.

Johnson and Chase (1981) subdivided Skinner's classification system for the purposes of instruction. Verbal tasks were categorized according to five of Skinner's functional classes (see Table 1). Practical use of these in achieving precise verbal instruction suggested a subdivision of Skinner's classes. For example, there appeared to be three kinds of intraverbals: stating definitions, providing original examples, and identifying written examples. Examples fitting these subclasses were selected from the literature of programmed instruction (Markle, 1967), concept programming (Miller & Weaver, 1976) and prose learning (Andre, 1979; Merrill & Tennyson, 1977: Rickards, 1979). Each type of task was defined in terms of the general relations between questions asked and answers given. For example, Skinner (1957) defined intraverbal as meeting these criteria: The antecedent is verbal; the response is verbal; and the response does not have either formal or point-to-point correspondence with the antecedent. Thus, a *definition task* is considered an intraverbal. If students are asked to "define reinforcement" and answer by describing the principle of reinforcement in the absence of a specific definition, then the relation fits the definition of an intraverbal.

This typology provides a basis for framing specific experimental questions. Two such questions were addressed in the present study: whether the different subclasses of the intraverbal have distinct, characteristic patterns of acquisition; and whether transfer of learning occurs across classes. If the training transfers from one type of task to another, then the tasks are functionally similar; if not, they are functionally different. The central issue was whether there are sufficient differences among various intraverbal relations to support a subclassification.

Three types of intraverbals were investigated: definition tasks, example identification tasks, and exemplification tasks (see Table 1 for examples). The questions asked were: (1) Does training on each of these intraverbal tasks result in differential rates of behavior during training? (2) Does training on each of these intraverbal tasks result in accurate performance on other intraverbal tasks during testing? (3) Does training on each intraverbal task result in differential responding on an intraverbal task that was never trained—that is, combination tasks?

METHOD

Subjects

Subjects were 6 undergraduate students (4 females and 2 males), selected from 12 who originally volunteered for the experiment. Three subjects were eliminated from the study because, during training, they did not exceed

Table 1
Typology of Verbal Instructional Tasks
Adapted from Johnson and Chase (1981).

Types	*Examples*
Echoic	Correctly repeat the following lines from Shakespeare's *Hamlet*. Be sure to copy my intonation closely.
Textual	Correctly pronounce the following (written) medical terms:
Transcriptive	
Copying	Correctly copy the following Chinese letters:
Dictation	Correctly spell the following names for laboratory equipment as I say them:
Intraverbal	
Definition	Define reinforcement.
Example Identification	Say which of the following written scenarios is an example of positive reinforcement:
Exemplification	Give an example of reinforcement.
Tact	
Example Description	Describe the technical properties of the plant specimens on the laboratory test table.
Example Identification	Say whether each of the following videotaped scenarios illustrates assertive or aggressive behavior:
Example Component Analysis	Identify at least three distinctive features of each of the wines in the goblets in front of you.
Combinations	Say whether this example illustrates reinforcement. If so, identify each defining feature. If not, identify the features that are present and change the example so that it illustrates reinforcement.

a correct response criterion of 75% in all conditions. Three subjects were eliminated because of experimenter errors in implementing the procedures. All subjects were sophomores and juniors majoring in psychology and had mastered before the experiment introductory-level concepts in both basic learning principles and experimental methodology. All subjects had failed to answer correctly the pretest questions concerning the concepts taught during the experiment.

Personnel

The first author (experimenter) coordinated the study; one male and one female undergraduate, both psychology majors, served as research assistants. The experimenter trained the assistants and checked the reliability with which they implemented the procedures. The training consisted of: (1) a detailed written description of the correct procedures for each session, (2) modeling of the procedures by the experimenter, (3) role playing, and (4) corrective feedback from the experimenter. Training on the concepts consisted of: (1) studying the prose passages for each concept, (2) answering all the questions or tasks, (3) feedback on performance, (4) reanswering all questions that had been answered incorrectly, and (5) terminal feedback. After training, the research assistants conducted the experimental sessions, corrected subjects' answers, and organized the data.

Setting

The study was conducted in two similar sound-insulated carrels. Each carrel was equipped with a desk, two chairs, and shelves for experimental materials. A one-way mirror connected the two carrels.

Response Class

The general intraverbal response class was defined as the presentation of a written question and the completion of a written answer that differed from the question. Completion was defined as the subject looking at the experimenter when he or she had finished writing an answer. The three training tasks and the four test tasks all conformed to this definition.

Materials and Apparatus

The experimental materials included three prose passages, each of which defined an esoteric psychological concept: abulia, constructional approach, and tau effect. For each concept a copy task, a set of example identification tasks, a series of definition tasks, a series of exemplification tasks, and two combination tasks were designed (see Table 1 for examples of these tasks).

Both the set of concepts and the specific tasks used for each concept were controlled for difficulty (Chase, 1980). The three concepts

were selected because they each met the following criteria:

1. They could be divided into five critical or defining features.
2. They could be defined in 125 to 150 words.
3. The experimenters could generate 25 or more different original examples of each.
4. A pilot test with 17 subjects revealed a similar level of difficulty for each.

The fourth criterion involved a selection test for concepts. Nine esoteric psychological concepts were defined and edited by the experimenters. Then each definition was given to 17 subjects. The subjects were asked to study the definition and complete a fill-in-the-blank question that defined the concept. Finally, when the subjects correctly defined the concept, they were asked to give three original examples of the concept.

"Abulia," "constructional approach," and "tau effect" were selected from the concept pool because subjects responded similarly on the tasks for these three concepts. First, all of the subjects defined these three concepts correctly. Second, 5 subjects wrote one correct original example of these concepts. Third, no subject wrote three original examples of these concepts.

The tasks used in the study were also pilot tested in order to eliminate tasks that were either too difficult or too easy for subjects and to help the experimenter decide how to distribute the tasks across the training and testing phases of the experiment. The training phases included a progression from easy to difficult tasks; the test phase included only the most difficult tasks from each type.

First, definition and exemplification tasks were written and edited by the experimenters. These tasks were designed to progress from defining and exemplifying each of the five features of each concept to synthesizing the five features into a complete definition or example.

Second, example identification tasks were written and edited by the experimenters; 25 examples and 25 nonexamples were written for each concept. The examples covered a range of situations in which the concept could be illustrated and each example illustrated the five features of the concept. The nonexamples also covered a range of situations in which the concept could be illustrated. However, each nonexample had one critical feature changed or deleted from the illustration. Thus, there were five nonexamples in which the first feature was modified, five nonexamples with the second feature modified, and so forth. As with the definition and exemplification tasks, the programming strategy was to make the critical features salient to the subjects (nonexamples) and to demonstrate how all the critical features were synthesized to form instances of the concept (examples).

Third, two combination tasks were constructed for each concept. Each combination task involved an illustration of a concept with the instructions: "Say whether this example illustrates the concept _____. If so, identify each of the defining features. If not, identify the features that are present and describe the changes that you would have to make in order to identify it as _____."

Fourth, 5 undergraduates were asked to answer all the tasks after reading the prose passages that defined the concepts. Explicit comments were obtained on three problem areas: communication problems (e.g., diction, grammar, sentence structure), motivation problems (e.g., boredom, ineffective attempts at humor), and step size (e.g., too many questions, too few questions). In addition, data on the accuracy and rate of correct responding were obtained.

These data revealed consistent responding on definition, exemplification, and combination tasks. All 5 students had similar response patterns on each type of task and changes were made on specific items on the basis of their comments. Items were ordered from those that asked for single features (the easiest items) to those that requested a complete synthesis of the features (the most difficult items). The complete definition and exemplification items as well as the combination items were used in the test phases of subsequent studies.

There was considerable variability on example identification tasks. Therefore, a second

pilot test was conducted for example identification tasks, after changes had been made in the items on the basis of the first subjects' comments. Twelve different undergraduate subjects were given a randomly ordered sequence of example identification tasks after they had read the prose passage defining a concept. They wrote "yes" for those illustrations that exemplified the concept, and "no" for those illustrations that did not exemplify the concept. To gain more information about possible misconceptions concerning the illustrations, subjects were also asked to orally justify their answers. This procedure was followed for each of the three concepts.

The results were analyzed in terms of a rank order of example identification tasks, calculated as the percentage of subjects who answered each task correctly. The percentages ranged from 0% to 100%. Then, the tasks were divided into five groups. Tasks that were answered correctly by fewer than 16% of the subjects were eliminated from the pool as being too difficult. Tasks that were answered by between 16% and 42% of the subjects were considered the most difficult and were reserved for the test phase of subsequent experiments. Tasks that were answered by between 50% and 66% of the subjects were considered moderately difficult and were used at the end of each study sequence. Tasks that were answered by between 75% and 92% of the subjects were considered easy and were programmed at the beginning of each study sequence. Finally, any task that was answered correctly by all the subjects was eliminated from the pool as being too easy. All of these procedures were adapted from Merrill and Tennyson (1977).

In addition to these materials, the study included a pretest, a scoring sheet, and procedural outlines for each experimental session. The assistants used the outlines to guide their conduct during the experiment. All instructional materials were typed and photocopied. Cassette tape recorders were used to record all interactions between assistants and subjects. The experimenter listened to the taped interactions to check the reliability with which the procedures were implemented. An electric timer was used to determine the duration spent by each subject on each task.

Procedure

The study was conducted with each subject individually, in four 1-hr sessions. The first session was devoted to assessing the subject's entering repertoires with respect to the concepts abulia, tau effect, and constructional approach. A pretest that included definition, exemplification, example identification, and combination tasks for each of the three concepts was administered. During the second through fourth sessions, similar general procedures were followed. Sessions differed only with respect to the type of intraverbal task used for training. These differences are specified in Table 2.

During each session, the general format was as follows: First, the subject read a prose passage that defined a concept. Second, the subject was asked to fill in the blanks of the copy task word-for-word, from the passage. (This was done to ensure that subjects actually read the material.) On completion, the assistant immediately corrected the copy task. If there were any mistakes, the subject was asked to correct them. Next, the series of tasks was presented for the specific condition that was assigned for the session. Specific detailed feedback, based on a prepared answer key, was given for each answer. If the subject's answer was correct, it was read to the subject. If the subject's answer was incorrect, the subject was told why—for example, "I'm sorry, that is not the right answer. Notice that abulia involves a sudden or abrupt change in reinforcement density, not just a change." Each task was timed separately. These procedures were followed until the subject completed the study sequence. Next, the test for that particular concept was given. The test consisted of nine examples and nonexamples, two terminal definition questions, two exemplification questions, and two combination questions (see Table 3 for details). The test was administered in the same way as the study sequence, except that no feedback was given.

After the session, the test was scored by an assistant. A detailed answer key was used that

Table 2
Description of Three Training Conditions Regardless of Concept

Define Condition	*Exemplification Condition*	*Example Identification Condition*
A. Prose passage defining the concept	A. Prose passage defining the concept	A. Prose passage defining the concept
B. Copy task as observing response	B. Copy task as observing response	B. Copy task as observing response
C. Define task on Feature 1 with prose passage	C. Exemplify task on Feature 1 with prose passage	C. Example of concept (with passage)
D. Define task on Feature 1[a] without prose[b] passage	D. Exemplify task on Feature 1 without passage	D. Nonexample of concept that varies Feature 1 (with passage)
E. Define task on Feature 2 with passage	E. Exemplify task on Features 1 & 2 with prose passage	E. Nonexample of concept that varies Feature 2 (with passage)
F. Define task on Feature 2 without passage	F. Exemplify task on Features 1 & 2 without passage	F. Nonexamples of concept that varies Feature 3 (with passage)
.	.	.
.	.	.
.	.	.
Terminal Define task[c] (complete definition) without passage	Terminal Exemplify task (complete original example) without passage	Three examples and nonexamples of concept without passage

[a]Tasks on the same feature are parallel, not identical.
[b]Exactly half the questions were answered with the passage available for referral by the subject. The other half were answered without the passage.
[c]The exact number of tasks in each condition varied as a function of concept. Constructional approach conditions usually required more tasks than abulia and tau effect.

analyzed each question in terms of the critical features or characteristics of each concept. Each prose definition was divided into its five critical features for the purpose of designing

Table 3
Description of the transfer tests for each concept (constructional approach, abulia, and tau effect) regardless of training condition[a].

A. Nine example identification tasks; one nonexample for each feature, four examples
B. Two define tasks
C. Two exemplify tasks
D. Two combination tasks; each broken into five parts

[a]All tests were scored with the following method:
1. Each example or nonexample was worth 1 point.
2. Each definition and exemplification task was worth 5 points and was scored according to the presence or absence of the 5 critical features for each concept.
3. Each combination task was worth 5 points and was scored according to whether the subject justified the presence of all the critical features or added or changed those that were absent or different.

Table 4
Critical feature analyses used for determining correct answers and correct classification of items.

All constructional approach items required:
1. Observation or interview to determine problem behavior.
2. Identification of reinforcing consequences of problem behavior.
3. Teaching an alternative behavior.
4. Consequating alternative with same reinforcers.
5. Keeping the environment the same.

All abulia items required:
1. High rate of behavior at first.
2. High ratio of reinforcer to behavior at first.
3. Ratio of reinforcer decreases.
4. Rate of behavior decreases.
5. Decrement in reinforcer/behavior ratio is abrupt.

All tau items required:
1. Three successive presentations of an identical event.
2. Time between 1st and 2nd is short.
3. Time between 2nd and 3rd is long.
4. Subject says the events are identical when time is short.
5. Subject says the events are different when time is long.

Table 5
Sequence of Training Conditions for Each Subject

Subjects	Session 2[a] Abulia	Session 3 Tau effect	Session 4 Con. app.
1, 5	Exemplify	Definition	Example ID
2, 7	Definition	Example ID	Exemplify
4, 8	Example ID	Exemplify	Definition

[a]Session 1 for each subject was the pretest.

an answer key. Table 4 presents the feature analyses of the three concepts.

The identification task required that all five features within the illustration be identified, the exemplification task required illustration of all five features, and the definition required description of all five features. In a combination task, the subjects had to describe those features that were present in the problem and illustrate those features that were absent. Although the five features were arbitrarily chosen, the feature analysis made the scoring of the different types of items comparable.

Experimental Design

An intrasubject, repeated measures design was used. Table 5 shows the sequence of the conditions for each subject. Each subject was trained with each of the study programs, studying each of the three concepts: constructional approach, tau effect, and abulia. All subjects were randomly assigned to one of three orders of training, according to a Latin-square counterbalanced sequence, to control for any order-of-treatment effects.

Reliability and Interscorer Agreement

The assistants' reliable implementation of the experimental procedures was ensured as follows: First, the experimenter listened to approximately 75% of the tapes and indicated any discrepancies with respect to the protocols. Second, the experimenter rescored approximately 45% of all the answers to study questions, calculating indices of interscorer agreement. (Low agreement indices would have indicated that subjects had received incorrect feedback. Therefore, their training would not have corresponded to the study condition to which they were assigned.) Table 6 presents the agreement data for this test. On the whole, the assistants implemented the study procedures as planned.

Table 6
Interscorer agreement indices. Each index calculated by dividing the number of agreements by the total number of agreements and disagreements.

	Mean	Median	Range
Training Questions	92.25%	94.00%	66-100%
Test Questions			
Example Identification	100.00%	100.00%	
Definition	93.37%	100.00%	66-100%
Exemplify	89.60%	100.00%	66-100%
Combination	88.30%	93.00%	70-100%
Durations	89.43%	86.00%	80-100%

Interscorer agreement indices for each dependent measure are also shown in Table 6. The experimenter rescored 25% of all tests to determine agreement, using the same answer keys and scoring procedures as those of the assistants. In addition, the experimenter observed 20% of the sessions through a one-way mirror and recorded the times that subjects spent completing tasks. The experimenter's and the assistants' records were compared and durations that differed by less than ± 2 s were considered agreements.

RESULTS

Training Performance

Figure 1 presents the acquisition data (e.g., performance during training) for the 6 subjects. The subjects are paired according to the combination of training program and concept that they received. All 6 subjects had higher correct rates answering example identifications than answering either of the other types of questions. In addition, example identification questions were answered less accurately than either definition or exemplifications. Five subjects answered definition questions more accurately than example identifications and 5 subjects answered exemplify questions more accurately than example identifications.

Group analyses were conducted for all dependent measures to determine whether variables such as concept, order of training, or the interaction between order and concepts signif-

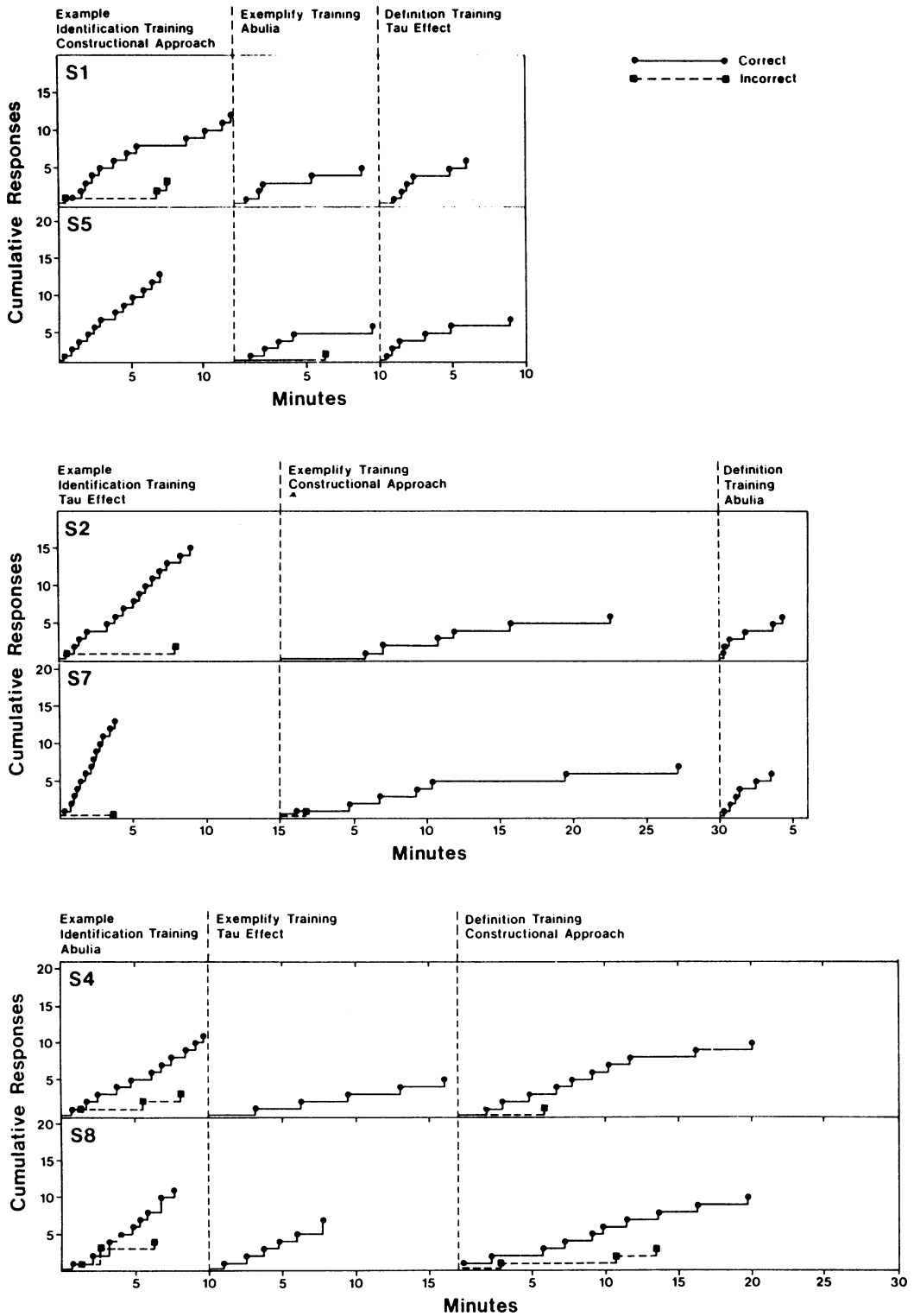

Fig. 1. Cumulative numbers of correct (circles) and incorrect (squares) responses as functions of time, for all 6 subjects on training tasks. The subjects are paired according to the combination of concept and training they received.

Table 7

Rate of correct responses per minute on items similar to those used in training (extension).

EXTENSION PERFORMANCE			
	Training Conditions		
Subjects	Example Identification	Exemplification	Definition
1	1.16	.38	.29
2	1.49	.10	.65
4	.86	.57	.22
5	.99	.60	.43
7	2.45	.21	.87
8	.66	.56	.35
Mean	1.27	.40	.47
Median	1.08	.47	.39

Table 8

Percent-correct performance on items similar to those used in training (extension) and items that were novel to the subjects (transfer).

EXTENSION PERFORMANCE			
	Training Conditions		
Subjects	Example Identification	Exemplification	Definition
1	89	80	80
2	100	40	60
4	56	100	20
5	78	100	80
7	90	100	80
8	56	100	80
Mean	78	87	67
Median	84	100	80

TRANSFER PERFORMANCE			
	Training Conditions		
Subjects	Example Identification	Exemplification	Definition
1	77	38	47
2	73	29	58
4	56	74	03
5	66	41	66
7	70	61	65
8	100	62	41
Mean	74	51	47
Median	71	51	52

icantly affected study performance. A three-way repeated measure, Latin-square ANOVA was calculated for each dependent measure. Order of training was the Latin-square factor, concept was the repeated factor, and training was the within-square factor. Each factor had three levels, and data from all subjects were used for the analyses. There were no significant concept effects, order effects, nor interactions on either rate or accuracy measures of acquistion.

Test Performance

Rate. Table 7 presents the rate of correct responding on test items. For all 6 subjects, rates of correctly answering extension tasks after example identification training were higher than rates of correctly answering extension tasks after definition training and exemplification training.

Accuracy. Table 8 presents the percent-correct performance (accuracy) on test items. Three critical differences were found among conditions. First, for most subjects accuracy on extension tasks was higher than transfer accuracy. Sixteen of the 18 intrasubject comparisons between extension and transfer revealed this difference. In addition, 13 of the 18 percent-correct scores on extension tasks were greater than 75% (the criterion for acquisition during training). However, only 2 out of 18 transfer scores were above 75%. Second, accuracy on extension tasks after exemplification training was higher than extension accuracy after example identification or definition training for 4 subjects. In addition, example identification extension performance was more accurate than definition extension performance for 4 subjects. Third, 5 subjects showed greater accuracy on transfer tasks after example identification than after either exemplification or definition training.

Combination tasks. Table 9 presents the subjects' performance on combination tasks. Because performance on the combination item was the only dependent measure that was identical across conditions, the effect of each training condition on combination performance was analyzed separately. Three differences were found among the conditions. First, the rate of correct responding was differentially affected by training. Five subjects had higher rates of correct combination responses after example identification training than after exemplification training, and 4 sub-

Table 9
Performance on Combination Items

RATE OF CORRECT RESPONSES			
	Training Conditions		
Subjects	*Example Identification*	*Exemplification*	*Definition*
1	.19	.31	.15
2	.38	.13	.24
4	.39	.33	.00
5	.36	.00	.30
7	.31	.16	.65
8	.24	.15	.36
Mean	.32	.18	.28
Median	.34	.15	.27

PERCENT CORRECT			
	Training Conditions		
Subjects	*Example Identification*	*Exemplification*	*Definition*
1	60	20	40
2	60	10	60
4	50	70	00
5	40	00	50
7	40	35	60
8	40	40	70
Mean	48	29	47
Median	45	28	55

jects had higher rates after example identification training than after definition training. Four subjects also had higher rates after definition than after exemplification training. Second, accuracy was differentially affected by training. Five subjects performed more accurately after definition training than after exemplification training, and 4 subjects, more accurately after example identification training than after exemplification training. Third, comparisons between combination performance and extension performance (compare Table 9 to Tables 7 and 8) revealed that most subjects performed more effectively on extension tasks. Sixteen of the 18 intrasubject comparisons on rate measures and 17 of the 18 intrasubject comparisons on accuracy measures showed this difference.

Control comparisons. As with the training performance, group analyses were conducted for measures of test performance. These analyses revealed a significant main effect of concepts on rate of correctly answering extension tasks, $F(2, 12) = 7.15$, $p < .025$. A planned comparison between concepts yielded a significant difference between constructional approach and the other concepts, $F(1, 12) = 12.12$, $p < .01$. In addition, analyses of rate revealed a significant interaction between order of training and concepts, $F(2, 12) = 4.50$, $p < .05$. Analyses of percent-correct performance also revealed a significant effect of concept, $F(2, 12) = 4.15$, $p < .05$. A planned comparison between constructional approach and other concepts was also significant, $F(1, 12) = 4.49$, $p < .05$. However, neither order effects nor interaction effects were found for accuracy measures of test performance.

DISCUSSION

These results both provide information and give rise to important questions concerning verbal behavior in general and verbal instruction in particular. First, the findings were consistent with previous claims that verbal behavior can be classified according to functional criteria (Bostow, 1976; Johnson & Chase, 1981; Skinner, 1957). Second, the functional differences among types of tasks and the functional similarity within types of tasks have practical implications for studying and teaching verbal behavior. Third, some of the control analyses revealed significant differences among concepts. Although this result is not surprising, it suggests the need for further refinements in a working functional classification system of verbal behavior. Each of these points is discussed below.

Both rate and accuracy data during training revealed that example identifications were substantially different from definitions and exemplifications. During acquisition, example identification was characterized by rapid, errorful responding, whereas exemplification and definition were characterized by slow, accurate performance. These results by themselves did not demonstrate the functional differences among the three intraverbal tasks. Other comparisons were required to conclusively establish the functional independence of these tasks, for topographical differences in response components of the particular tasks

used here could account for the differences found during training. However, given the care that was taken to make these responses as topographically similar as possible and the assumption that similar rates of acquisition should be obtained from similar operants regardless of topography, we included these data in the analysis. Although the effect of a response can be defined as a momentary change in the environment, measuring a series or patterns of changes in behavior over time may be a more thorough, though indirect, measure of effect.

The test data revealed further differences. First, complete transfer of learning was not obtained from one type of intraverbal to other intraverbals. Subjects responded less accurately on those tasks that were not trained. Second, most of the differences in rate and accuracy found during acquisition were maintained during the tests: Exemplifications were slow and accurate, example identifications were rapid and errorful, and definitions were slow. In addition, new differences emerged from the test data. Definitions were not answered as accurately as exemplifications, transfer accuracy was differentially affected by the type of training, and performance on combination items was differentially affected by the type of training.

In general, most comparisons indicated systematic differences between example identification and other types of performance. Half of the comparisons revealed differences between exemplification and definition performance. Finally, raising the probability of each type of intraverbal did not result in similar increases in the other intraverbals. Therefore, these results support the claim that within programmed verbal learning, the intraverbal can be subdivided into functionally distinct classes of verbal behavior (Johnson & Chase, 1981).

Basic studies of verbal behavior might benefit from a classification system that extends from broad functional categories like the intraverbal to more specific classes like those described here. It may be possible to standardize the operants used across studies and generally to assist communication among investigators. Similarly, this would help instructors to teach and evaluate different classes of verbal behavior.

Example identification training may be considered the most efficient because it took the least amount of time, but total test accuracy (transfer and extension) was equally high after other kinds of training. These findings were consistent with previous research on the effects of examples and nonexamples on concept acquisition (Keenan & Grant, 1979; Miller & Weaver, 1976; Watts & Anderson, 1971).

The low degree of transfer between types of intraverbals indicates that it is not sufficient to teach and evaluate one kind of intraverbal if one wants students to engage in a variety of intraverbals. Future research should address the training conditions that lead to transfer to a variety of types of intraverbals. Perhaps models such as the stimulus equivalence model (Sidman & Tailby, 1982) provide some insight into the kinds of training conditions that will lead to high levels of transfer across types of intraverbals.

These findings are complicated, however, by the differences found among concepts on test performance. These differences should not be regarded merely as a source of variability that limited the conclusions about functional similarity of types of intraverbal tasks. The concept differences found on the test indicate that verbal learning involves complex interactions among stimulus events and behavior. Thus, definitions of verbal learning may require classifying more than the general task relation. That is, descriptions and predictions of verbal relations may become more accurate when task classifications are combined with a classification scheme for conceptual stimuli. Typically, the operant definition of a concept includes both conceptual stimuli and conceptual behavior (Keller & Schoenfeld, 1950). The present study, however, concentrated on classifying the conceptual behavior and did not classify the differences between conceptual stimuli that involved one term—for example, constructional approach—versus stimuli that involved another term—say, abulia.

The data suggest that operant analyses of verbal behavior will be facilitated by defining very specific verbal operants, for subclasses of

intraverbal tasks were found to have characteristic patterns of responding. The effect of verbal behavior could be predicted through classifying it according to the relationship between questions that ask for a particular type of response and the students' responses. More precise control may be possible if future studies analyze the combined effects of the type of response requested and the type of conceptual stimulus involved in the request.

REFERENCES

Andre, T. (1979). Does answering higher-level questions while reading facilitate productive learning? *Review of Educational Research,* **49,** 280-318.

Azrin, N. H., Holz, W., Ulrich, R., & Goldiamond, I. (1961). The control of the content of conversation through reinforcement. *Journal of the Experimental Analysis of Behavior,* **4,** 25-30.

Baron, A., Kaufman, A., & Stauber, K. A. (1969). Effects of instructions and reinforcement-feedback on human operant behavior maintained by fixed-interval reinforcement. *Journal of the Experimental Analysis of Behavior,* **12,** 701-712.

Boe, R., & Winokur, S. (1978). A procedure for studying echoic control in verbal behavior. *Journal of the Experimental Analysis of Behavior,* **30,** 213-217.

Bostow, D. E. (1976). *The experimental analysis of college instructional practices: Some mid-course corrections.* Invited address presented at the second annual convention of the Midwestern Association of Behavior Analysis, Chicago.

Catania, A. C., Matthews, B. A., & Shimoff, E. (1982). Instructed versus shaped human verbal behavior: Interactions with nonverbal responding. *Journal of the Experimental Analysis of Behavior,* **38,** 233-248.

Chase, P. N. (1980). *A methodology for evaluating classification schemes and verbal learning variables.* Unpublished master's thesis, University of Massachusetts, Amherst.

Chomsky, N. (1959). A review of B. F. Skinner's *Verbal Behavior. Language,* **35,** 26-58.

Frisch, S. A., & Schumaker, J. B. (1974). Training ing generalized receptive prepositions in retarded children. *Journal of Applied Behavior Analysis,* **7,** 611-621.

Galizio, M. (1979). Contingency-shaped and rule-governed behavior: Instructional control of human loss avoidance. *Journal of the Experimental Analysis of Behavior,* **31,** 53-70.

Greenspoon, J. (1955). The reinforcing effect of two spoken sounds on the frequency of two responses. *American Journal of Psychology,* **68,** 409-416.

Guess, D., Sailor, W., Rutherford, G., & Baer, D. M. (1968). An experimental analysis of linguistic development: The productive use of the plural morpheme. *Journal of Applied Behavior Analysis,* **1,** 297-306.

Hart, B. M., & Risley, T. R. (1968). Establishing use of descriptive adjectives in the spontaneous speech of disadvantaged preschool children. *Journal of Applied Behavior Analysis,* **1,** 109-120.

Harzem, P., Lowe, C. F., & Bagshaw, M. (1978). Verbal control in human operant behavior. *Psychological Record,* **28,** 405-423.

Holz, W. C., & Azrin, N. H. (1966). Conditioning human verbal behavior. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application* (pp. 790-826). New York: Appleton-Century-Crofts.

Johnson, K. R., & Chase, P. N. (1981). Behavior analysis in instructional design: A functional typology of verbal tasks. *Behavior Analyst,* **4,** 103-121.

Keenan, J. B., & Grant, L. (1979). *The effects of definitions, examples and nonexamples, and feedback on student conceptual responding.* Paper presented at the Fifth Annual Conference on Personalized Instruction, Washington, DC.

Keller, F. S., & Schoenfeld, W. N. (1950). *Principles of psychology.* New York: Appleton-Century-Crofts.

Lamarre, J., & Holland, J. G. (1985). The functional independence of mands and tacts. *Journal of the Experimental Analysis of Behavior,* **43,** 5-19.

Lee, V. L. (1978). Teaching generalized receptive and productive behind-front discriminations to two retarded children. *Journal of Applied Behavior Analysis,* **11,** 529.

Lee, V. L. (1981a). Preposition phrases spoken and heard. *Journal of the Experimental Analysis of Behavior,* **35,** 227-242.

Lee, V. L. (1981b). Terminological and conceptual revision in the experimental analysis of language development: Why. *Behaviorism,* **9,** 25-53.

MacCorquodale, K. (1970). On Chomsky's review of Skinner's *Verbal Behavior. Journal of the Experimental Analysis of Behavior,* **13,** 83-99.

Markle, S. M. (1967). Empirical testing of programs. In P. Lange (Ed.), *Programmed instruction* (pp. 104-138) Chicago: University of Chicago Press.

Merrill, M. D., & Tennyson, R. D. (1977). *Teaching concepts: An instructional design guide.* Englewood Cliffs, NJ: Educational Technology Publications.

Miller, L. K., & Weaver, F. H. (1976). A behavioral technology for producing concept formation in university students. *Journal of Applied Behavior Analysis,* **9,** 289-300.

Osgood, C. E. (1949). The similarity paradox in human learning: A resolution. *Psychological Review,* **56,** 132-143.

Place, U. T. (1981a). Skinner's *Verbal Behavior:* I. Why we need it. *Behaviorism,* **9,** 1-24.

Place, U. T. (1981b). Skinner's *Verbal Behavior:* II. What is wrong with it. *Behaviorism,* **9,** 131-152.

Place, U. T. (1982). Skinner's *Verbal Behavior:* III. How to improve Parts I and II. *Behaviorism,* **10,** 117-136.

Rickards, J. P. (1979). Adjunct postquestions in text: A critical review of methods and processes. *Review of Educational Research,* **49,** 181-196.

Sailor, W. (1971). Reinforcement and generalization of productive plural allomorphs in two retarded children. *Journal of Applied Behavior Analysis,* **4,** 305-310.

Segal, E. (1977). Toward a coherent psychology of language. In W. K. Honig & J. E. R. Staddon (Eds.), *Handbook of operant behavior* (pp. 628-653).Englewood Cliffs, NJ: Prentice-Hall.

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expan-

sion of the testing paradigm. *Journal of the Experimental Analysis of Behavior,* **37,** 5-22.
Skinner, B. F. (1957). *Verbal behavior.* New York: Appleton-Century-Crofts.
Taylor, C. (1972). *Peaceful coexistence in psychology.* Paper presented at the 80th Annual Convention of the American Psychological Association, Honolulu.
Watts, G. H., & Anderson, R. C. (1971). Effects of three types of inserted questions on learning from prose. *Journal of Educational Psychology,* **62,** 387-394.
Wheeler, A. J., & Sulzer, B. (1970). Operant training and generalization of a verbal response form in a speech-deficient child. *Journal of Applied Behavior Analysis,* **3,** 139-147.

Received May 1, 1984
Final acceptance March 18, 1985