DOCUMENT RESUME

ED 075 990 EM 010 905

AUTHOR	Tennyson, Robert D.
TITLE	Effect of Negative Instances in Concept Acquisition Using a Verbal Learning Task. Tech Memo Number 71.
INSTITUTION	Florida State Univ., Tallahassee. Computer-Assisted Instruction Center.
REPORT NO	FSU-CAI-TM-71
PUB DATE	30 Jan 73
NOTE	35p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Adverbs; Comparative Analysis; *Concept Formation; Generalization; Grade 7; *Grammar; *Teaching Methods; *Verbal Learning
IDENTIFIERS	Exemplars; Nonexemplars

ABSTRACT

An experiment replicated and extended a previous study on the value of negative instances in concept learning (i.e., an instance of something which does not exemplify the concept helps to make clear what the concept is). In this experiment, the concept to be learned was "adverb". The subjects were seventh grade students. Some groups of subjects read sentences in which the adverb was identified. Other groups of subjects read pairs of sentences; in one of these sentences, the word was used as an adverb and was identified as such, and in the other sentence, the same word was used as a different part of speech. The different function the word served in each sentence of the pair was apparent. Results showed that subjects learned the concept of adverb better when the negative example was available. (JK)

CAI CENTER

ED 075990

TECH MEMO

EFFECT OF NEGATIVE INSTANCES IN CONCEPT ACQUISITION
USING A VERBAL LEARNING TASK

Robert D. Tennyson

Tech Memo No. 71
January 30, 1973

Center for CAI
The Florida State University
Tallahassee, Florida 32306

EM 010905

FLORIDA STATE UNIVERSITY

Tech Memo Series

The FSU-CAI Center Tech Memo Series is intended to provide communication to other colleagues and interested professionals who are actively utilizing computers in their research. The rationale for the Tech Memo Series is three-fold. First, pilot studies that show great promise and will eventuate in research reports can be given a quick distribution. Secondly, speeches given at professional meetings can be distributed for broad review and reaction. Third, the Tech Memo Series provides for distribution of pre-publication copies of research and implementation studies that after proper technical review will ultimately be found in professional journals.

In terms of substance, these reports will be concise, descriptive, and exploratory in nature. While cast within a CAI research model, a number of the reports will deal with technical implementation topics related to computers and their language or operating systems. Thus, we here at FSU trust this Tech Memo Series will serve a useful service and communication for other workers in the area of computers and education. Any comments to the authors can be forwarded via the Florida State University CAI Center.

Duncan N. Hansen
Director
CAI Center

EFFECT OF NEGATIVE INSTANCES IN CONCEPT ACQUISITION USING A VERBAL LEARNING TASK

Robert D. Tennyson

Tech Memo No. 71
January 30, 1973

Center for CAIR
The Florida State University
Tallahassee, Florida 32306

Effect of Negative Instances in Concept Acquisition Using a Verbal Learning Task

Robert D. Tennyson

Florida State University

ABSTRACT

Concept acquisition was promoted by manipulating positive and negative instances. Also, stimulus similarity variables produced the concept classification errors of overgeneralization, undergeneralization, and misconception. The value of negative instances in concept instruction was investigated in a second adverb learning task by removing the negative instances from the six treatment conditions. Subjects (260 total) were seventh grade students from three school districts. Subjects in experiment one responded according to the hypothesized outcomes ($p < .01$). In experiment two, the subjects responded randomly on the posttest. The results indicate that negative instances are an integral part of concept acquisition. The relationship between the positive and negative instances was based upon similarity of irrelevant attributes and sentence difficulty.

Effect of Negative Instances in Concept Acquisition
Using a Verbal Learning Task

Robert D. Tennyson
Florida State University

The earliest research dealing with the relationship of negative instances (nonexemplars) to positive instances (exemplars) was Smoke's (1933). Smoke used an artificial task in which the exemplars and non-exemplars were randomly ordered with the order changing after each succession through the list. No logical relationship was established between exemplars and nonexemplars, resulting in his conclusion that negative instances were of no value in concept learning. A study which looked at the relationship of exemplars based on critical attributes was Morrisett and Hovland's (1959) replication of Adams' (1954) study of single task vs. multiple task. They found that a variety of positive instances was necessary to effect a transfer of concept attainment. However, there was no operational definition of the relationship between exemplars according to their irrelevant attributes or any criteria. In investigations of combined instances, the equivalent attributes of positive and negative instances were found to be poorly utilized by human subjects (Bruner, Goodnow, & Austin, 1956; Donaldson, 1959; Hovland & Weiss, 1953). These studies show the lack of an operational control between exemplars and nonexemplars, as does the Smoke study. The concepts used in these prior studies were finite with the subject attempting to guess the rule (critical attribute) from a series of instances.

Callentine and Warren (1955) studied positive instances and concluded that repetition of one or two instances increased attainment. Luborsky (1945) indicated that eight exposures was more effective than three. These last two studies show that a series of instances is needed, but no mention is made of the difficulty of the instances or that discrimination of negative instances could be affected by a series which includes a combination of exemplars and nonexemplars. Irrelevant attributes as measures of difficulty has been shown in the studies dealing with ease of attainment of concept classes (Archer, Bourne, & Brown, 1956; Brown & Archer, 1956; and Bourne, 1957). Each of these studies found that as the number of irrelevant attributes increased the learning latency and number of errors also increased. They concluded that the number of irrelevant attributes has a linear relationship with difficulty of instances.

Negative instances facilitate learning of concepts by requiring the subject to concentrate on the critical attribute when presented a matched relationship of exemplars and nonexemplars. When the exemplar and nonexemplar are as similar as possible in their irrelevant attributes, the noticeable differences are among the critical attributes. Concept attainment research does not provide the subject with the opportunity to focus on the critical attributes by using negative instances. Without negative instances in concept acquisition, the subject might conceive as a critical attribute an irrelevant attribute.

In a study by Tennyson, Woolley, and Merrill (1972), independent variables were investigated that predicted concept acquisition and specified classification errors. The results of their study were based on three stimulus similarity variables: probability, matching, and divergency. The probability variable referred to the difficulty of the instances. Probability of each instance was a percentage of subjects who correctly identified it given only a definition. The matching variable refers to the relationship between exemplars and nonexemplars. A matched condition was defined as instances having similar irrelevant attributes. Divergency referred to the relationship between two exemplars. Exemplars were divergent when their irrelevant attributes were as different as possible. By logically manipulating the three independent variables into four treatment conditions, Tennyson et al., predicted four dependent variables. They were (1) Correct Classification, all instances, exemplars and nonexemplars, correctly identified; (2) Overgeneralization, nonexemplars similar to class members identified as exemplars; (3) Undergeneralization, low probability exemplars identified as nonexemplars; and (4) Misconception, exemplars and nonexemplars sharing a common irrelevant attribute identifed as class members. The four strategies consisted of presenting to subjects a definition and task according to the hypotheses: (1) IF high to low probability, divergent, and matched, THEN correct classification; (2) IF low probability, divergent, and not matching, THEN overgeneralization; (3) IF high probability, divergent, and matching, THEN undergeneralization; and (4) IF high to low probability, convergent, and not matching, THEN misconception.

Hypotheses

The purposes of the present investigation were: (a) to replicate the Tennyson et al. (1972) study with a different population and task to add external validity to the findings; (b) to extend the Tennyson et al. study in terms of manipulating the stimulus similarity variables to see the effect of matching in the undergeneralization dependent variable and probability rating in the overgeneralization dependent variable (Table 1); and (c) the effect of negative instances in concept acquisition.

The study was divided into two experiments for clarification of discussion. Experiment One was the replication of the Tennyson et al. (1972) study with the addition of two classification error conditions (Table 1). The null hypotheses were predicted for Experiment Two. Negative instances were removed from each of the six conditions hypothesized in Experiment One (Table 1). The alternate hypotheses indicated that the same dependent variables would result even with the exclusion of the non-exemplars (Table 3). In each condition the null hypotheses predicted a random response pattern from the subjects. The removal of the negative instances would eliminate any observable learning of the concept.

Method

Learning Task

The behavioral objective of the task was: Given a sentence, the subject will identify by circling, any adverb(s) in the sentence. Concept acquisition is required because the subject is presented exemplars and nonexemplars in instruction and then required to generalize to previously unencountered exemplars on the posttest, as well as discriminating unencountered nonexemplars. The adverb concept was chosen for three reasons.

TABLE 1

Hypotheses Matrix

Dependent Variable Outcomes	Independent Variables Presented			
	Probability	Matching	Divergency	Nonexemplars*
1) Correct Classification	All Levels	Matched	Divergent	Yes
2) Over-generalization	Low Levels	Unmatched	Divergent	Yes
3) Under-generalization	High Level	Matched	Divergent	Yes
4) Misconception	All Levels	Unmatched	Convergent	Yes
5) Over-generalization A	All Levels	Unmatched	Divergent	Yes
6) Under-generalization A	High	Unmatched	Divergent	Yes

* In Experiment Two the nonexemplars were removed from the six treatment conditions.

First: the concept provides an unlimited number of instances. A sentence arrangement would never appear twice. Second: the concept is generally taught throughout the United States' schools. It seemed relevant to use a classroom concept that is part of the English language program. Third: students would learn verbs, nouns, and adjectives prior to the Christmas holiday. A short review would follow the holiday and then adverbs would be introduced, followed by pronouns, prepositions, etc. The students had the preparation of understanding verbs but still were unacquainted with adverbs as presented in formal education. This presentation was timed to be the subjects' first introduction to adverbs. The definition of the adverb used in the learning task was:

An adverb is a word that modifies (changes meaning of) a verb, an adjective, or another adverb and answers the question: When? How? Where? How much? or To what extent? See the following examples:

The plane flew yesterday (The adverb modifies the verb.)

The dinner was not good. (Modifies the adjective.)

She sang very well. (Modifies another adverb.)

When?	Where?	How?	How much?	To what extent?
once	outside	lamely	rather often	runs faster
today	here	silently	very poor	dances best

Selection of sentences used in the study followed the procedures outlined by Tennyson and Boutwell (1972). First, a subjective selection was made by identifying the critical attributes of the definition, and the more common irrelevant attributes. After 120 sentences were developed in sets of four--two divergent examples matched to two nonexemplars--an empirical analysis was conducted. The sentences were randomly scrambled and divided in two equal parts. This provided two lists which were then renumbered in reverse order making four tests. A definition of adverbs was given on the first page of the test along with directions requiring the subject to circle the adverbs in the sentences that followed. The results were ordered in terms of frequency of correct responses along a continuum ranging from 10% to 84%. The mean score was 52%. High probability sentences were those correctly defined by the subjects at a frequency of 65% level and above; low probability sentences were identified at 35% level and below.

Procedure

The programs for the six treatments in Experiment One (Table 1) followed the same format display: general directions, pretest on verbs, pretest on adverbs (half the subjects), task, and posttest. Upon concluding the general directions, read by the experimenter while subjects read silently, the subjects turned to page one and began the self-instructional program. A pretest was taken by all subjects to determine level of knowledge of verbs. The pretest, used as a covariate, consisted of 20 sentences. The subject was directed to circle the verb in each sentence. Half the subjects were then given a second pretest on adverbs. This test required subjects to identify adverbs in a sentence by circling them. Following the pretest(s) a definition and brief explanation of adverbs was presented to all subjects except those in the control group. The only instruction for the rest of the program was the presentation of 16 sentences. The format of the exemplar/nonexemplar displays consisted of eight pairs of sentences--two each per page, for example:

Example:	She is never home. (verb)
Not an Example:	Never is a long time. (noun)
Example:	His papers are too messy. (adj.)
Not an Example:	His two papers are messy. (noun)

The exemplars and nonexemplars were identified along with underlining the adverb and diagraming the modifying word (prompting variable). The word modified by the adverb was identified according to its part of speech. There were no adverbs in the nonexemplar sentences.

When finished with the program, all subjects took the posttest. Subjects were directed to read each sentence and identify the adverb by circling it. There were 30 sentences arranged in order on two pages. When finished with the posttest, the subject turned over the program on the desk and left the testing room.

Experiment Two programs were identical to those in Experiment One, except that the nonexemplars were removed. Subjects received the same pretest on verbs. Half the subjects took the adverb pretest and all the subjects read the definition of adverbs. Subjects in Experiment Two received the posttest when finished with the program (same as Experiment One subjects). An example of the Experiment Two format per page is as follows:

Slowly she walked home.
(verb)

Debbie is <u>sound</u> asleep.
(verb)

She is <u>never</u> home.
(verb)

His papers are <u>too</u> messy.
(adj.)

Sixteen selections of poetry were selected as the nonrelevant task for the control group. These subjects took the pretest on verbs, and half took the pretest on adverbs followed by a definition of trochaic meter and directions to study the poetry. When completed with their task on poetry, the control group took the same posttest as the experimental groups.

The programs were printed and stapled together in a self-instructional booklet. Once the student began, no questions concerning the program were answered by the experimenter. Directions requested the subjects to not return to previous pages. Since the program was nonspeeded, the subject could spend as much time per page as desired. The posttest was attached at the end of the program. The subjects were directed to leave quietly when finished since it was anticipated that subjects receiving Experiment Two would finish before others.

<u>Treatment Programs</u>

The programs for the twelve treatment conditions are identified by the dependent variable names. For the correct classification program, the exemplar sets were arranged from high to low probability. Page one had a high probability set of divergent exemplars with matched nonexemplars followed on page two with a high medium set, page three a low medium set, and page four with a low probability set. This task was hypothesized to result in a subject generalizing to all exemplars on the adverbs post-test and discriminating the nonexemplars by not identifying them as exemplars. A sample of the first page of sentences for the classification program shows high probability exemplars and nonexemplars:

Divergent

Example: (Matched) Slowly she walked home. (verb)

Not an Example: (Matched) She had a slow walk home. (noun)

Example: (Matched) Debbie is sound asleep. (verb)

Not an Example: (Matched) Debbie is a sound sleeper. (noun)

The overgeneralization program was constructed of only low probability exemplar sets, that is, all four pages of the program were low probability exemplars. However, the nonexemplars from the exemplar sets were changed. The new nonexemplars were randomly chosen from other sets. Such an unmatched situation would prevent the subject from distinguishing the critical attribute of adverbs from some other form of grammar. This situation is the condition found in concept attainment research in which the subject sees no relevancy in the negative instances. And as a result, the subject fails to develop a discrimination strategy. Contrast the first page of the overgeneralization program (below) with the first page of the classification program.

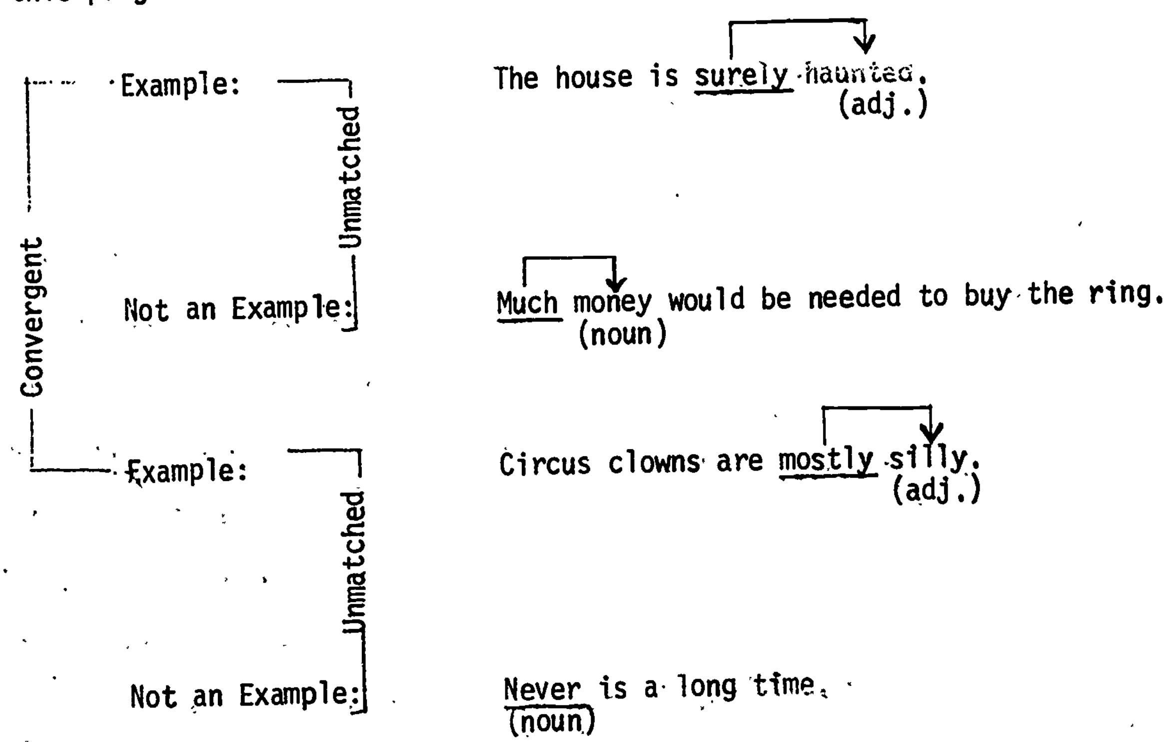
Divergent	Example:	Unmatched	It was not difficult to explain the word.
	Not an Example:		She had a slow walk home. (noun)
	Example:	Unmatched	The merchant sold out his candy supply. (verb)
	Not an Example:		The pretty baby smiled. (noun)

The undergeneralization task was constructed of only high probability exemplar sets. Since the nonexemplars were matched, the entire set was used. The first page of this program was the same as the correct classification program, with the succeeding pages on an equal level of difficulty. The example shown here is the last page:

Divergent	Example:	Matched	The book was easily understood. (verb)
	Not an Example:		It was an easy book to understand. (noun)
	Example:	Matched	I shall try harder. (verb)
	Not an Example:		I missed the harder problem. (noun)

It was hypothesized that by using only exemplars that had probability ratings of 60% or above, subjects would not generalize to previously unencountered low probability instances because they had not seen the difficult irrelevant attributes.

For the misconception program the convergent grouping was the "ly" ending; exemplar sets ended in "ly." The exemplar sets included high and low probability ratings. Nonexemplars randomly picked from other exemplar sets replaced the matched nonexemplars. Following is an example from this program:



With the use of only "ly" ending adverbs and with unmatched nonexemplars that did not have "ly" endings, it was hypothesized that the subject receiving this condition would respond to the irrelevant attribute of "ly" endings (whether or not connected to adverbs).

The second overgeneralization (A) program (Table 1) differed from the first in that a range of probability exemplar sets was used. Exemplar sets were the same as used in the correct classification program but without the matching of nonexemplars. The discrimination problem was hypothesized to occur because of the presentation of unmatched negative instances.

The only change for the second undergeneralization (A) program (Table 1) was the use of unmatched nonexemplars in the high probability exemplar sets. The matched nonexemplars were removed and replaced with a random selection of unmatched items.

Tests

The pretest on verbs was a standardized test taken from the Teacher's Manual of The New Building Better English - Grade Seven. The pretest on adverbs was also taken from that text. Construction of the posttest (Table 2) was based upon the same criteria as discussed in the treatment programs section. The test items were taken from the same pool of sentences as the treatment programs, but were not used in instruction or on the pretest. They were previously unencountered instances. Test construction followed this outline:

1. Convergent high probability exemplar.
2. Convergent low probability exemplar.
3. High probability nonexemplar matched to number 1.
4. Low probability nonexemplar matched to number 2.
5. High probability nonexemplar unmatched.
6. Divergent high probability exemplar paired to number 1.
7. Divergent low probability exemplar paired to number 2.

8. High probability nonexemplar matched to number 6.
9. Low probability nonexemplar matched to number 7.
10. Low probability nonexemplar unmatched.

The purpose of the above system is to predict subject responses according to type of exemplars and nonexemplars used in instruction. Any number of the above sets can be included in a posttest. For this study, three sets were used. The 30 sentences were randomly scrambled so that no patterns were evident to the subjects. To test the dependent variable of misconception, the words ending in "ly" were identified as convergent, all other endings were classified as divergent. Subjects in the misconception treatment condition were hypothesized to classify only convergent high and low probability exemplars and identify as exemplars those matched nonexemplars. The classification treatment group was hypothesized to correctly classify all exemplars on the test. The overgeneralization treatment group was hypothesized to classify not only the exemplars but the low probability matched nonexemplars as instances. They could also pick high probability matched and unmatched nonexemplars and not be penalized by an error. A subject in this group could have classified all 30 items as exemplars and still follow the predicted results. Undergeneralization subjects were hypothesized to respond only to the high probability exemplars.

The hypotheses were stated so that the resulting condition would have fewer errors when its scoring pattern was used against the other groups (Table 3). For example, the correct classification group would have zero errors, while the overgeneralization group would have eight errors, the undergeneralization group having the second fewest errors (six), the misconception group having the most (nine), the second overgeneralization group with eight, and the second undergeneralization group with six. For each of the other dependent variables, when scored with its pattern, the error would be zero.

The hypothesized response patterns for each of the dependent variables are given in Table 2. A subject was scored with an error for a given dependent variable when his response to a given item differed from the predicted response. Scores were obtained for the three sections of the test and then added together for the four separate dependent variable conditions. This procedure gave the subject four scores, one for each hypothesized dependent variable.

Experimental Design

A Solomon four design was used so that interaction of pretesting and the treatment could be analyzed to control external validity (Campbell and Stanley, 1963). For each treatment, one-half of the subjects received a pretest on adverbs. A three-way analysis of covariance was used to analyze the data. The three main effects were program treatments, pretest versus no pretest on adverbs, and sex. The covariate was the pretest on verbs. Test for homogeniety of regression of within-class and between-class linearity was performed.

In both experiments the control group subjects were the same. The three-way design increased the number of treatments to 26. Treatment programs were scrambled by a random number table (Li, 1964) and assigned to subjects. The basic experimental unit was the individual subject.

TABLE 2

Scoring Sheet

	SET #1	S	M	O	U	C
1.	eg #6		+	+	+	+
2.	eg #16		+	+	-	+
3.	$\overline{eg}$ #3		?	?	-	-
4.	$\overline{eg}$ #30		+	+	-	-
5.	$\overline{eg}$ #15		?	?	-	-
6.	eg #12		-	+	+	+
7.	eg #21		-	+	-	+
8.	$\overline{eg}$ #8		-	?	-	-
9.	$\overline{eg}$ #17		-	+	-	-
10.	$\overline{eg}$ #4		-	+	-	-

Note.--Predicted responses according to conditions. M - misconception; O - overgeneralization; U - undergeneralization; C = correct classification; + = S indicates this sentence is a positive instance; - = S indicates this sentence is a negative instance; ? = S could classify as either, no error possible; eg indicates an exemplar; $\overline{eg}$ indicates a nonexemplar; # refers to original test item number.

Statistical power was calculated from Cohen's (1969) book, with a moderate F test effect size (ES) of .25 for the three main effects. An alpha of .05 was used for all analysis of covariance tests. The total N-size was 260.

Subjects

The instance probability analysis was conducted with 110 seventh grade students from Springville Middle School (Utah). Three school districts, Alpine, Provo, and Nebo, provided the 260 subjects used in the study. The number of subjects randomly selected from each school was Lincoln Middle School (Alpine)--87 subjects out of 354 total seventh grade enrollment; Farrer Middle School (Provo)--68 subjects out of 275 total seventh grade enrollment; Dixon Middle School (Provo)--29 subjects out of 286 total seventh grade enrollment; Payson Middle School (Nebo)--27 subjects out of 247 total seventh grade enrollment; Spanish Fork Middle School (Nebo)--25 subjects out of 269 total seventh grade enrollment; and Springville Middle School--24 subjects out of 274 total seventh grade enrollment. No subjects were dropped from the investigation.

Results

Variable Measures

Each subject had four independent error scores obtained from responses on the posttest (Table 2). Table 3 shows the treatment groups, represented by capital letters, and the predicted errors for each dependent variable; that is, the C (correct classification) group would make zero errors under the correct classification variable but it was predicted that O (overgeneralization) group would make eight errors, the U (under-generalization) group six errors, the M (misconception) group would make nine

TABLE 3

Hypothesized Error Responses and Adjusted Mean Error Scores

Dependent Variables	Experiment 1								Experiment 2				Control
Groups	C	O	U	M	OA	UA	CN	ON	UN	MN	OAN	UAN	
Class	6.1*	12.3	10.2	11.5	12.9	10.4	11.9	13.3	11.8	11.6	11.9	12.9	13.0
	0**	8	6	9	8	6	0	8	6	9	8	6	
Over.	9.6	7.5	13.0	11.9	7.7	13.0	12.3	12.7	11.5	13.4	10.1	14.8	13.8
	8	0	14	11	0	14	8	0	14	1	0	14	
Unde.	9.2	13.4	7.1	10.9	12.9	7.1	9.9	11.1	10.4	9.4	11.8	10.8	11.3
	6	14	0	9	14	0	6	14	0	9	14	0	
Misc.	10.2	11.4	10.0	6.5	12.0	10.2	9.9	9.7	10.5	10.0	10.0	10.6	10.0
	9	1	9	0	1	9	9	11	9	0	11	9	

Note.--The treatment groups are represented by capital letters: C = Correct Classification; O = Overgeneralization; U = Undergeneralization; M = Misconception; OA = Overgeneralization A; UA = Undergeneralization A. Treatments ending with N are those which did not receive nonexemplars.

*First rows are the adjusted mean scores.

**Second rows are the predicted error scores.

errors, and OA and UA groups making the same as the O and U respectively. The alternate hypotheses for Experiment Two predicted that the treatment groups would make the same errors as in Experiment One. Thus, each group was predicted to make significantly fewer errors than the other conditions when its own dependent variable was analyzed. Likewise, the other variations in error scores per group were predicted.

The adjusted covariate means for the 13 groups according to the dependent variables are listed in Table 3. A separate three-way analysis of covariance was used for each dependent variable. The four covariate F tests (df = 12,208, α = .05, power = .75) on the treatment main effect were: Correct Classification, F = 8.35; Overgeneralization, F = 10.42; Undergeneralization, F = 9.05; Misconception, F = 6.03. The main effect of pretest on adverbs versus no pretest was nonsignificant (df = 1,208, α =.05, power = .97), except on the Overgeneralization analysis of covariance. The pretest group had an adjusted mean of 12.25, while the no-pretest group had a mean of 11.11. Since the pretested group had the higher mean it was assumed that the pretest had no effect on the posttest results. Sex was nonsignificant (df = 1,208, α = .05, power = .97), except on the Overgeneralization analysis. The boys had a lower adjusted mean error score of 11.28 to 12.08 for the girls. If this main effect had been significantly consistent with the other analyses, a detailed study would have been conducted. It was assumed that sex was not a factor in this study. No interactions were significant (p > .05). The posteriori tests used were the Newman-Keuls Sequential Test and Duncan's New Multiple Range Test.

Experiment One

Correct Classification. Subjects receiving the correct classification program were predicted to identify all the adverbs on the posttest without responding to other words as exemplars; that is, they would make zero errors on the posttest. The other conditions were hypothesized to make significantly more errors (Table 3). On both the Newman-Keuls and Duncan's NMRT, the C group made fewer errors than the O, M, U, OA, and UA groups and the control group ($p < .01$). This corresponds to the hypothesis and the predicted responses in Table 3. There was a difference between the control group and the U and UA groups on Duncan's NMRT ($p < .05$). The O and U groups differed on Duncan's NMRT ($p < .05$). According to Table 3, there was a predicted difference of two errors between the U group and the O group. A three-point error spread was predicted between U and M groups but none resulted ($p > .05$).

Overgeneralization. Two overgeneralization conditions were investigated to analyze the probability variable. In condition one only low probability instances were used, while in condition two a range was used. The two overgeneralization groups were nonsignificant in number of errors ($p > .05$) on the posttest (Table 3). The Newman-Keuls showed a difference between the two overgeneralization groups and the two undergeneralization groups ($p < .01$); this follows the prediction from Table 3 of an error spread of 14 points. A difference existed between overgeneralization groups and M and control groups on both posteriori tests ($p < .01$). Only on the Duncan's NMRT was there a difference between overgeneralizations groups and C group ($p < .05$). Other predicted error differences from Table 3 on both tests were between undergeneralization groups and C group; and the control group and overgeneralization and C groups ($p < .01$). On Duncan's NMRT there was a difference between C group and M group ($p < .05$).

Undergeneralization. Two undergeneralization conditions were investigated, both having high probability exemplars, but one would have matched nonexemplars and the other unmatched. The mean errors for the two undergeneralization groups were nonsignificantly different($p < .05$). The multiple comparisons of the undergeneralization error scores show for both the Newman-Keuls and Duncan's NMRT that the undergeneralization groups differed from the two overgeneralization groups ($p < .01$), a difference of 14 points was predicted. The hypothesized mean difference between the undergeneralization groups and the M group (11 points) and control group resulted ($p < .01$). The undergeneralization groups and C group error means were not similar for the Newman-Keuls ($p < .05$) and Duncan's NMRT ($p < .01$); the predicted errors between the C group undergeneralization groups was six (Table 3). Other predicted differences were: O group and C group on both tests ($p < .01$); and O group and M group on Newman-Keuls ($p < .05$) and Duncan's NMRT ($p < .01$). There were no error differences between O group and M group (five point difference was predicted) ($p > .05$).

Misconception. A misconception error was hypothesized if subjects were presented convergent exemplars, similar irrelevant attributes, and unmatched nonexemplars. In this investigation the irrelevant attribute in which the subject was hypothesized to accept as relevant was the "ly" ending. The results followed the predicted variables on all factors on both tests ($p < .01$), that is, the M group differed from the O, OA, U, UA and control groups. No mean differences resulted from the comparison on the other five groups as predicted in Table 3 ($p > .05$).

Experiment Two

The treatment conditions in Experiment Two were the same as Experiment One except for the removal of nonexemplars. Table 3 shows that the predicted response errors for each group were the same as Experiment One.

Correct classification. The paradigm for concept acquisition assumes a relationship between positive and negative instances. In concept attainment the subject is given a set of objects which keep reoccurring until mastery. As a result, negative instances are not utilized by the subject. However, concept acquisition mastery is determined by identification of new instances. The subject in this treatment is predicted not to discriminate or even generalize without negative instances in the instruction because of the limited number of positive instances presented. These subjects received the same program as subjects in the correct classification program of Experiment One; that is, divergent exemplars on a range of probability, but without the nonexemplars. Removal of the nonexemplars was predicted to produce random responding on the posttest. On both multiple comparison tests the error means in Experiment Two were nonsignificant between the groups ($p > .05$). The C group in Experiment One made the fewest errors on the posttest. However, no error difference resulted between the CN group and the control group ($p > .05$).

Overgeneralization. The discrimination problem of unmatched nonexemplars is further aggravated by the complete removal of negative instances. With unmatched nonexemplars the subject was aware of some differences between the two, but without nonexemplars the subject was not sure of the relationship between the critical and irrelevant attributes. In a program of low probability exemplars which were divergent and with no nonexemplars, random responses by the subject were hypothesized. The ON group and the OAN group had similar errors on Newman-Keuls ($p > .05$), but there was a difference on Duncan's NMRT ($p < .05$). The OAN group differed from the UAN group and the control group on both tests ($p < .01$). OAN group also deviated from MN and CN groups on both tests ($p < .01$). The UN and UAN groups were different on Newman-Keuls ($p < .05$) and Duncan's NMRT ($p < .01$). An error change resulted between the O and OA groups (Experiment 1) and the ON group on both tests ($p < .01$). The difference between OAN group and O and OA groups was on Duncan's NMRT ($p < .05$).

Undergeneralization. Generalizing to new instances was hypothesized to result from using divergent exemplars. Negative instances, by their presence, focused the subject's attention to the irrelevant attributes that aided in the ability to generalize as well as discriminate. It was predicted that subjects receiving only high probability divergent exemplars, without nonexemplars, would randomly respond on the adverb posttest. Groups UN and UAN had the same number of errors ($p > .05$). There were no mean variations between UN and UAN groups and the other groups ($p > .05$). MN group differed from OAN group on Duncan's NMRT ($p < .05$). A difference between the U and UA groups (Experiment 1) and UN and UAN groups was on Duncan's NMRT ($p < .01$), and Newman-Keuls with U and UA groups and UAN group ($p < .01$), and UN group ($p < .05$).

Misconception. Misconception in concept acquisition was the result of the subject receiving exemplars which stressed the same irrelevant attributes. Matching exemplars and nonexemplars tended to correct this problem. Presenting convergent exemplars without nonexemplars was hypothesized to result in random responses. The convergent irrelevant attribute would not produce the same effect as the matched instruction in Experiment One's misconception condition because the subject would not see the irrelevant attribute as a nonexemplar. Thus, the assumption that the subject would respond to that irrelevant attribute when tested on previously unencountered instances was not hypothesized. On the contrary, random responses were predicted because the subject would not distinguish the relevant attribute. On the adverb posttest subjects were predicted to identify instances at random. There were no differences between the groups on this dependent variable ($p > .05$). The M group (Experiment 1) had fewer errors than the MN group ($p < .01$).

Discussion

The results of the two experiments added validity to the Tennyson, Woolley, and Merrill (1972) study because the conclusions of the data were similar. Experiment One was a direct replication of that earlier study with the extension of the overgeneralization treatment which manipulated independent variables of probability, and undergeneralization treatment which manipulated matched exemplars. The first overgeneralization group had the condition of low probability exemplars only, with the second overgeneralization group receiving a range of probability. The two undergeneralization treatment groups reversed the variable condition of matched and unmatched. The study was extended to seventh graders rather than college students. Experiment Two introduced the fourth variable of exemplar/nonexemplar presentation to analyze the effects of negative instances on concept acquisition.

The independent variable, divergency, dealing with the relationship between exemplars according to their irrelevant attributes was significant. The misconception group, instructed with adverbs ending with the same irrelevant attribute, identified as relevant the nonexemplars with that same irrelevant attribute. The other treatment conditions received divergent exemplars and did not respond to the irrelevant attribute when associated with a nonexemplar. Generalization within a concept class was a function of the divergency variable. By instructing with very different exemplars, subjects transferred more readily when tested with previously unencountered exemplars. The correct classification and overgeneralization groups received divergent exemplars and responded to the more difficult exemplars on the adverb posttest. The difficulty of exemplars was determined by the instance probability analysis which subjectively and empirically rated the instances according to ease of recognition.

Probability, as an independent variable, is unique because instances can be rated on difficulty prior to developing instruction. Sequencing of easy-to-difficult instances can make instruction more attuned to the individual. Subjective rating of items has been the usual procedure in all forms of instructional development. The instance probability analysis is a heuristic approach to defining the levels of difficulty of instances. As irrelevant attributes intensity, the difficulty of the instance increases. Research as early as Bourne's (1957) has shown this linear relationship. Obtaining a subjective analysis on each instance (see Tennyson & Boutwell, 1972) and then constructing matched pairs with nonexemplars is the first step in deciding which instances to use in the instruction of a concept class. By combining the subjective analysis with a probability rating, the sequencing of exemplar sets eliminates much of the guesswork in program development. The most significant difference obtained in Experiment One was between the two undergeneralization groups and the two overgeneralization groups. A generalization problem was hypothesized when the subject received only divergent high probability instances. The subject would not transfer to low probability adverbs on the posttest. The subject receiving this treatment made fewer responses on the posttest than any other group. It was hypothesized that the overgeneralization problem, where the subject would not discriminate previously unencountered exemplars from nonexempl would be promoted by using divergent low probability instances of adverbs. Subjects in this treatment condition responded not only to exemplars but to large numbers of nonexemplars. They identified more words as adverbs than any other group. The use of the probability variable alone did not cause this problem. Nonexemplars used in the instruction were unmatched with exemplars. In the second overgeneralization condition, the subjects received a range of probability on exemplars. This would indicate that the overgeneralization error is an interaction effect of probability. This conclusion is supported by the fact that the correct classification group was instructed with exemplar sets that had matching and with sets sequenced in a range of probability that resulted in the fewest number of errors of any group.

The effect of the matching variable was shown by the increased response to nonexemplars by the two overgeneralization groups on the adverb posttest. In both cases, the nonexemplars were unmatched to the exemplars so that subjects failed to recognize the critical attributes from the irrelevant attributes. When given difficult exemplars, the subjects did respond to the adverbs on the posttest. They could generalize to new adverb instances, but they could not discriminate from words that were not adverbs. The misconception group had an unmatched relationship between exemplars and nonexemplars and, likewise, the subjects failed to distinguish between the critical and irrelevant attributes of the adverbs on the posttest. The result of the response patterns on the posttest show that the misconception group responded frequently to words ending in the irrelevant attribute of "ly," while not responding to other irrelevant attributes as the two overgeneralization groups did. Thus, by using convergent exemplar sets the subjects focused on a common irrelevant attribute shared by all exemplars and assumed that to be relevant. The interaction effect of matching with the other variables is illustrated by the two overgeneralization groups and the correct classification group. One undergeneralization group received a matched situation while the other situation was unmatched. There was no significant difference between their scores. Therefore, probability was a more effective variable in producing an undergeneralization problem on the posttest. The correct classification group was instructed with a matched relationship of exemplars and nonexemplars and resulted in a significantly lower error score on the adverb posttest than the undergeneralization groups with matched nonexemplars and the overgeneralization group with a range of probability--the two treatment conditions studied by the correct classification group.

All of the treatment conditions in Experiment One had nonexemplars. The subjects responded on the adverb posttest as hypothesized. Experiment Two directions and instructions were the same as Experiment One except for the removal of the nonexemplars. The alternate hypotheses predicted the same response patterns for the subjects on the posttest as in Experiment One. The null hypotheses for Experiment Two were supported in this study because of the instructional paradigm which hypothesized that concept. Acquisition discrimination is taught by exemplar sets that include matched nonexemplars having irrelevant attributes as similar as possible. On the adverb task used in this study, the subjects in Experiment Two responded randomly on the posttest. The interpretation is that the subjects completely failed to acquire the concept of adverbs when presented just positive instances. The failure was both a generalization and discrimination problem. With the removal of the negative instances from the correct classification program for the CN condition in Experiment Two, the subjects did no better than the control group. There were fewer significant differences in Experiment Two between error means, and when these differences did occur they were not according to the alternate hypotheses response patterns as in Experiment One. Concept acquisition even with divergent exemplars and probability is incomplete without negative instances.

Correct classification is a result of the interaction of the four independent variables investigated in this study. A variance which cannot be explained is that the error mean score of the correct classification group, while significantly lower than the other conditions, was still high. An extension of this study to account for the variance should include sequencing exemplar sets on a more individual basis. A subject might require more or fewer high probability exemplars based upon personality differences. Other extensions could include individual probability ratings done by a heuristic program on a CAI terminal, a more precise measure of the matching variable, and defining both relevant and irrelevant attributes.

REFERENCES

Adams, J. A. Multiple versus single problem training in human problem solving. Journal of Experimental Psychology, 1954, 48, 15-18.

Archer, E. J., Bourne, K. E., Jr., & Brown, F. G. Concept identification as a function of irrelevant information and instructions. Journal of Experimental Psychology, 1956, 52, 316-321.

Bourne, L. E., Jr. Effects of delay of information feedback and task complexity on the identification of concepts. Journal of Experimental Psychology, 1957, 54, 201-207.

Brown, F. G., & Archer, E. J. Concept identification as a function of task complexity and distribution of practice. Journal of Experimental Psychology, 1956, 52, 316-321.

Bruner, J. S., Goodnow, J., & Austin, G. A. A study of thinking. New York: Wiley, 1956.

Callentine, M. F., & Warren, J. M. Learning sets in human concept formation. Psychological Review, 1955, 1, 363-367.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1963.

Donaldson, J. Positive and negative information in matching problems. British Journal of Psychology, 1959, 50, 253-262.

Hovland, C. I., & Weiss, W. Transmission of information concerning concepts through positive and negative instances. Journal of Experimental Psychology, 1953, 45, 175-182.

Li, J. C. Statistical inference. Ann Arbor, Michigan: Edwards Brothers, Inc., 1964.

Luborsky, L. F. Aircraft recognition: I. the relative efficiency of teaching procedures. Journal of Applied Psychology, 1945, 29, 385-398.

Markle, S. M., & Tiemann, P. W. Really understanding concepts. Champaign, Ill.: Stipes, 1969.

Mechner, F. Science education and behavioral technology. In R. Glaser (Ed.), Teaching machines and programed learning, II: Data and directions. Washington, D.C.: National Education Association, 1965.

Merrill, M. D. Necessary psychological conditions for defining instructional outcomes. In M. Merrill (Ed.), Readings in instructional design. Englewood Cliffs, N. J.: Prentice-Hall, 1971.

Morrisett, L., Jr., & Hovland, C. I. A comparison of three varieties of training in human problem solving. Journal of Experimental Psychology, 1959, 58, 52-55.

Smoke, K. L. An objective study of concept formation. Psychology Monographs, 1933, 42 (4, Whole No. 191).

Tennyson, R. D., & Boutwell, R. C. A quality control design and evaluation model for hierarchical sequencing of programed instruction. NSPI Journal, 1971, 10, 5-10.

Tennyson, R. D., & Merrill, M. D. Hierarchical models in the development of a theory of instruction: A comparison of Bloom, Gagné, and Merrill. Educational Technology, 1971, September, 27-30.

Tennyson, R. D., & Woolley, F. R. The effect of interaction of anxiety with performance on two levels of task difficulty. Journal of Educational Psychology, 1972, 62, 463-467.

Tennyson, R. D., Woolley, F. R., & Merrill, M. D. Exemplar and non-exemplar variables which produce correct concept classification behavior and specified classification errors. Journal of Educational Psychology, 1972, in press.

Woolley, F. R., & Tennyson, R. D. A conceptual model of classification behavior. Educational Technology, 1972, in press.