

*THE MODIFICATION OF SENTENCE STRUCTURE AND ITS  
RELATIONSHIP TO SUBJECTIVE JUDGEMENTS  
OF CREATIVITY IN WRITING<sup>1</sup>*

KAREN BLASE MALONEY AND B. L. HOPKINS

UNIVERSITY OF KANSAS

The present study objectively defined and manipulated some compositional variables in 10-sentence stories written by fourth-, fifth-, and sixth-grade students, and related these operationally defined variables to subjective judgements of creativity. Points, exchangeable for candy and extra recess, were given to members of two teams contingent upon their using different adjectives, different action verbs, and different sentence beginnings. The students' use of these selected parts of speech was modified and the independent subjective ratings indicated that stories written during contingency conditions were generally rated as more creative than those written during baseline conditions. Operational definitions that specify non-repetition or variety of responses, and contingencies that require response diversity may provide a beginning basis for defining writing creativity and the conditions that maximize its occurrence.

Since the early 1950s, much has been written about creativity (Mooney and Razik, 1967). However, little of the work on creativity simultaneously relates to applied settings, deals with socially important behaviors, and employs adequate experimental analyses to identify controlling variables. Maltzman (1960) employed appropriate experimental controls to strengthen the "creative" behavior of college students, but the responses measured were arbitrary original word associations. Other studies (Torrance and Tan, 1964; Osburn, 1957) have dealt with more complex, socially relevant, and less arbitrary behavior but have not employed experimental controls to identify controlling variables. Much of the work with creativity has speculatively analyzed creative behaviors (Maslow, 1963; Gutman, 1961), or has attempted to find

correlations among creative behaviors and other behaviors, or between various attributes, personality characteristics, and test scores (MacKinnon, 1970; Flescher, 1963; Getzels and Jackson, 1961).

One problem that may pre-empt the experimental analyses of socially relevant creative behavior is the difficulty of developing reliable, operational definitions for those behaviors. Some of the response properties purported to be necessary for a behavior to be creative are unusualness, appropriateness, transformation of material to overcome conventional constraints, and condensation, which warrants close and repeated examination of the response (Jackson and Messick, 1965).

Although such a list of response properties is extremely useful, it is apparent that such properties require operational definitions in order to be utilized in an experimental analysis of creativity. The problem of defining creativity is further compounded by the fact that a given response may be judged as a creative response, relative to either the norms of the population of which the individual is a member, or relative to the given individual's past behavior. This latter aspect has been incorporated into the response definition of more recent applied research.

---

<sup>1</sup>This research was supported in part by the Educational Personnel Development Act, Training Teachers in Behavior Modification in Early Childhood Education, Grant No. OEG-0-70-1820(721). Special thanks to Miss Martha Smith, Miss Marion Fesmire, Miss Melanie Lusk, Miss Kathy Meyer, and Mrs. Barbara El-Antably for their assistance and cooperation. Reprints may be obtained from the first author, Western Carolina Center, Morgantown, N.C. 28655, or from the second author, Department of Human Development, University of Kansas, Lawrence, Kansas 66044.

Within the context of this vein of research, one functional operational definition of an aspect of creative behavior is the occurrence of a different response to the same stimulus situation. For example, Pryor, Haag, and O'Reilly (1969), reinforced only those gross motor responses of a porpoise that had never been previously observed in the training of that porpoise. Similarly, a recent study dealing with human behavior in an applied setting has been reported by Goetz and Baer (1971). In that study, social reinforcement, contingent upon blockbuilding forms not previously produced by a given child for a particular construction, increased form diversity. Another study, conducted by Goetz and Salmonson (1972), employed a similar procedure with preschool children to increase creative easel painting. In these studies, creativity is, in part, defined as, or in some way identified with, novel behavior not previously displayed in a specific setting or session.

The present study, employing a similar definition, represents an attempt to define and manipulate some compositional variables in stories written by fourth-, fifth-, and sixth-grade children, and to relate these operationally defined variables to subjective ratings of creativity.

## METHOD

### *Subjects and Setting*

Fourteen elementary school children, grades four through six, attending a voluntary, non-remedial six-week summer school session, participated in the study. All students occupied the same classroom at the same time. The classroom was under the supervision of a sixth-grade teacher who was assisted by two aides, one of whom was the experimenter.

### *General Procedures*

At the beginning of each class, a noun, randomly selected without replacement from a pool of 30 nouns, was written on the blackboard. Students were instructed to write a 10-sentence story, using the noun as the story topic.

Upon completion, or at the end of 40 min, each child handed his story to the experimenter, and went on with his assigned work.

The experimenter then scored each paper on compositional variables for which points were given. Ten minutes before morning recess, the number of points each child had earned was written besides his or her name on the blackboard.

### *Response Measures*

The students' 10-sentence stories were scored on a number of objective compositional variables by a scorer, who was not informed about the purposes of the experiment, or the experimental procedures employed. The compositional variables scored and the response definitions employed were:

*Number of letters.* Any of the 26 letters of the alphabet, either capitalized or not capitalized.

*Number of sentences.* A sentence was defined as beginning with a capital letter and/or on a new line and/or having a period, question mark, or exclamation point at the end containing at least one subject and predicate.

*Number of words.* Any group of letters, written or printed, representing a spoken word. Misspelled words were counted as words.

*Number of adjectives.* A word serving as modifier of a noun, to denote quality of the thing named, to indicate quantity or extent, or to specify a noun as distinct from something else.

*Number of adverbs.* A word that modified a verb, adjective, another adverb, preposition, phrase, clause, or a sentence, and that expresses some relation of manner of quality, time, place, degree, number, cause, opposition, affirmation, or denial.

*Number of action verbs.* A word that expresses an act, occurrence, or movement, but *not* a mode of being, as any form of the verb "to be".

*Number of prepositional phrases.* A group of words beginning with a preposition and combining with a noun, pronoun, or other noun

equivalent to form a phrase that typically has an adverbial, adjectival, or substantive relation to some other word.

*Number of compound sentences.* A sentence having two or more main clauses.

The number of sentences with more than eight words was also scored. The number of *different* adjectives, adverbs, and action verbs and the number of *different* sentence beginnings were also scored for each paper.

Different was defined as a word used for the first time in a given 10-sentence story; or when defining different sentence beginnings as a word used only once as the first word of a sentence. Once a word was scored in a given category, repetitions of that word could not be scored again for that given paper.

In addition to the scorer's counting each of the previously mentioned variables, the experimenter scored, in class, those variables for each condition, for each story, for which points were contingently given. These experimenter-scored variables were: number of different adjectives, Days 6 to 9; number of different action verbs, Days 10 to 13; and number of different sentence beginnings, different adjectives, and different action verbs, on Days 14 to 17. However, the independent scorer's data were always used when computing and plotting data. The number of minutes and seconds for each child to complete each assignment was also recorded by the experimenter.

### *Reliability*

Several types of reliability checks were made throughout the study. First, to record the number of minutes and seconds for each assignment to be completed, the experimenter started a stopwatch as soon as the last sheet of blank paper was handed out. The time, in minutes and seconds, was recorded at the top of each child's paper. An independent observer also read the running stopwatch from three to six times per condition, and recorded the time on a piece of paper, out of the view of the experimenter. Reliability of recording times varied from 0 to 5

sec disagreement with a mean disagreement of 2.1 sec.

Reliability measures were also taken for each of the previously mentioned objective, compositional variables, as follows: two scorers, one of whom scored all papers for the study, and neither of whom were given any information about the purposes of the experiment, independently scored Xeroxed copies of three children's papers for each of four conditions. Therefore, a total of 12 reliability checks for each variable was taken.

For the number of letters, words, and sentences, reliability was computed by dividing the smaller number by the larger. For all other measures, reliability was computed as  $100 \times$  the number of agreements divided by the number of agreements plus disagreements.

Mean reliability for number of letters was 97%; words, 94%; and sentences, 97%. For number of adjectives, the mean reliability was 89%; for different adjectives, 88%. Inter-scorer reliability was 81% for adverbs and 87% for different adverbs. Mean reliability scores for action verbs and different action verbs were 88% and 84%. On other variables, mean reliability was 98% for number of different beginnings, 78% for compound sentences, 89% for sentences with more than eight words, and 89% for prepositional phrases.

### *Experimental Conditions*

*Baseline.* During baseline conditions, Days 1 to 5, the general procedures were in effect. That is, 10-sentence stories were written after the topic noun was written on the board. Ten minutes before recess, 100 was written on the blackboard next to the name of each student that had handed in a 10-sentence story.

*Good Writing Game.* Throughout the remainder of the study, a Good Writing Game was played, which was modelled after the Good Behavior Game (Barrish, Saunders, and Wolf, 1969). For this game, the class was divided into two teams, which were listed on the blackboard at the back of the room. The experimenter had

previously determined the members of each team by rank-ordering the 14 students in pairs from the highest to the lowest pair, according to the mean number of points each student had accumulated over 10 of the objective compositional variables, across the five days of baseline. A point was given for each adjective, adverb, action verb, different adjective, and adverb and action verb, different sentence beginning, compound sentence, prepositional phrase, and sentence with more than eight words. These points were not at any time seen by the students and were used only by the experimenter in rank-ordering students. Members of rank-ordered pairs were randomly assigned to Team 1 or Team 2.

Each team member received a specified arbitrary number of points for using the experimenter-designated compositional variable without repetition. Each winning team member went to recess 5 min early and received a small piece of chocolate candy. Both teams could win if their total number of team points was greater than the number posted on the blackboard above their respective teams. This number was 80% of a total possible team maximum set by the experimenter. If both teams were below criterion, then both could still win when the difference between team totals was 100 points or less.

The explanation of the above-mentioned game procedures occurred only on Day 6. However, at the beginning of each session, a statement of the contingency in effect for earning points and three examples of the behavior required were written on the board. For instance, on Days 6 to 9, three examples of different adjectives were verbalized by students, then written on the board and verbalized by the experimenter. The examples were erased before the topic noun and a statement of the contingency "20 different adjectives = 100 points" were written on the board. This general procedure was employed each day with the compositional variable and contingency appropriate to the condition.

*Different adjectives.* For Days 6 to 9, five points were given for every different adjective used in a story; no points were given for repeating an adjective.

*Different action verbs.* Ten points were given for each different verb showing action in a given story on Days 10 to 13. No points were given for action verbs that were repeated in a given story.

*Different adjectives, different action verbs, and different beginnings.* Five points were given for every different adjective, 10 points for every different action verb, and 10 points for every sentence that began with a different word in a given story. No points were given for repeating an adjective, action verb, or sentence beginning in a story.

Pupil-teacher and pupil-experimenter interactions, in all conditions, were limited to answering questions about the spelling of words.

Reliability measures were taken on the compositional variables for which points were given in each condition. During reliability checks, the experimenter would list the words for which points were given and then compare that list with the scorer's list of the same objective compositional variables. Reliability was computed as  $100 \times$  the number of agreements divided by the number of agreements plus disagreements. Mean reliability for Days 6 to 9 for different adjectives was 88%; for different action verbs on Days 10 to 13 it was 90%. Reliability for Days 14 to 17 for number of different beginnings was 99%, for number of different adjectives, 86%, and for different action verbs, 75%. Mean reliability averaged 86% over Days 14 to 17.

Furthermore, two independent raters were employed in order to provide some comparison between the experimental results and subjective judgements of creativity. One rater was a graduate student in English and the other a graduate student in German. Raters were not informed of experimental procedures or results.

Each rater was given four compositions of each of the 14 students. Raters were instructed

to order each set of four compositions from most creative to least creative. They were also verbally instructed not to base their rankings on spelling, punctuation, handwriting, or the topic of the story, which had been assigned. Raters were further instructed only to compare the four compositions of one child with one another. That is, raters were instructed to make only intra-subject comparisons of creativity.

Both independent raters were given the same 14 sets of four compositions, but at different times. The four compositions per child were randomly selected, one from each of the four conditions. Reliability was 46%, that is, out of 54 papers the raters agreed exactly on 25 ratings. Chance reliability would be an agreement of 14 out of 54. A Chi-square analysis shows this to be significant at the 0.0001 level.

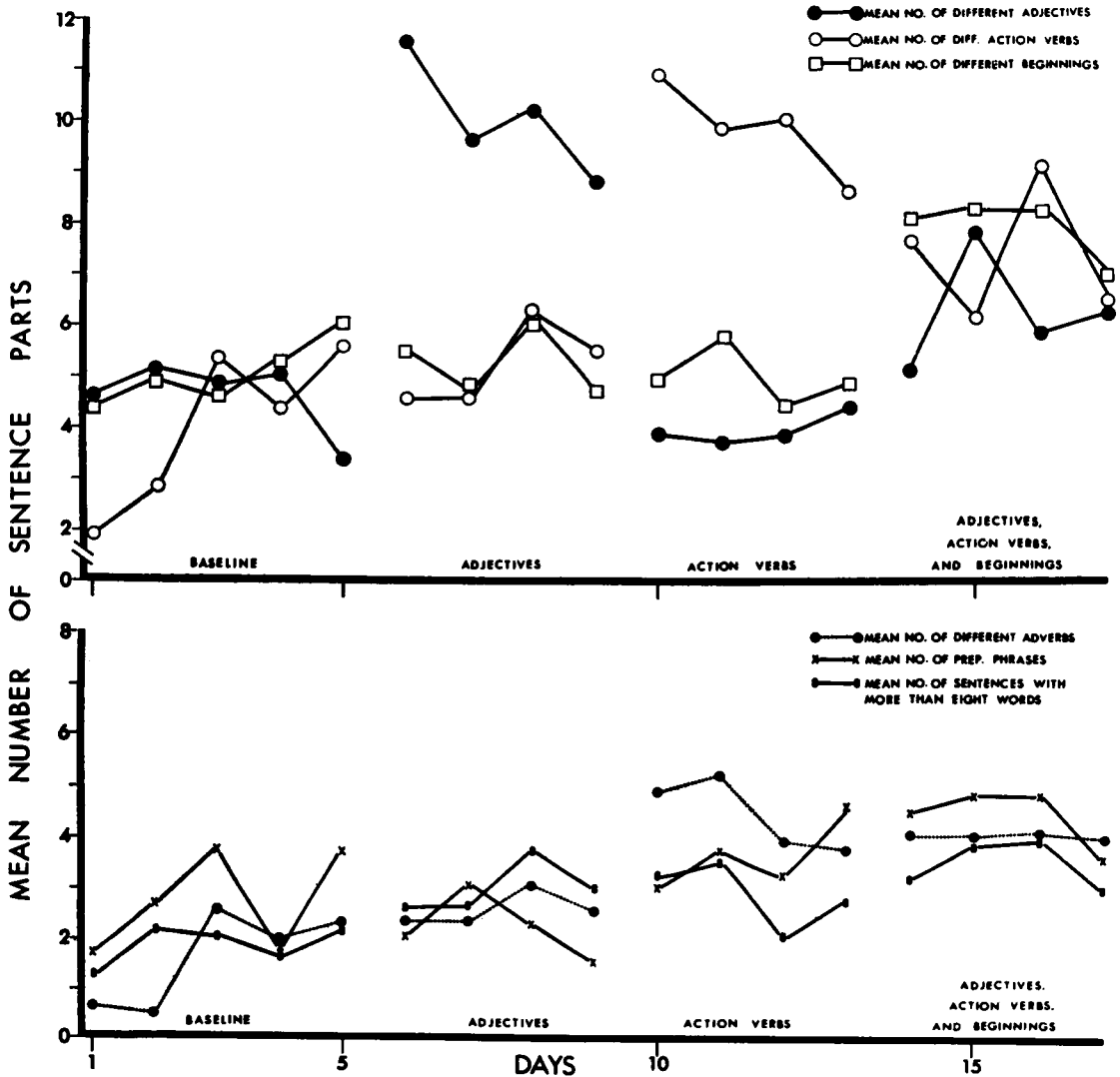


Fig. 1. (upper co-ordinates) Mean number of different adjectives, different action verbs, and different beginnings for 14 ten-sentence stories for Days 1 to 17.

Fig. 2. (lower co-ordinates) Mean number of different adverbs, prepositional phrases, and sentences with more than eight words for 14 ten-sentence stories for Days 1 to 17.

## RESULTS

Figure 1 on the upper co-ordinates illustrates the main experimental results. During baseline, the mean number of different adjectives was 4.59 across the condition. This increased to a condition mean of 9.95 during Days 6 to 9, when points were given for the use of different adjectives. The mean number of different action verbs and the mean number of different beginnings were 5.19 and 5.22 respectively on Days 6 to 9.

With the change in contingencies to points for different action verbs, on Days 10 to 13, the mean number of different adjectives dropped from 9.95 to about 3.9, while the condition mean for different action verbs increased to 9.75. Mean number of different sentence beginnings remained about the same.

However, when contingencies were placed on all three compositional variables on Days 14 to 17, the mean number of different beginnings increased to about 7.8. In this condition, the mean numbers of different adjectives and different action verbs were higher than in baseline or when the contingencies for the sentence part were not in effect, but were lower than when winning the game was contingent on number of different adjectives on Days 6 to 9, or on number of different action verbs on Days 10 to 13. These intermediate means were about 6.15 for different adjectives, and 7.19 for different action verbs.

Figure 2 on the lower co-ordinates illustrates the results for three other compositional variables: the mean number of different adverbs, the mean number of prepositional phrases, and the mean number of sentences with more than eight words. The mean number of different adverbs was lowest during baseline, 1.65, and increased very slightly to about 2.6 during the different adjective condition. The greatest increase in the number of different adverbs occurred when points were given for action verbs. Different adverbs were then at about 4.47 and decreased only slightly to 4.1 per 10-sentence story when

different adjectives, action verbs, and beginnings were reinforced.

The mean number of prepositional phrases was slightly higher in baseline than when points were given for different adjectives. The mean number of prepositional phrases then increased slightly during the last two conditions.

The mean number of sentences with more than eight words was lowest during baseline, but remained fairly constant across Days 6 to 17, with condition means of 3.03, 2.93, and 3.54.

Mean work rate, defined as mean number of letters and words written per minute, did not vary appreciably from condition to condition, nor did the mean number of compound sentences.

### *Individual Results*

Of the 14 students, eight students' results were highly similar to the class means in Figure 1. These eight students are represented by the condition means of one student in Figure 3.

Highly similar is defined as results that first, illustrate an increase over baseline in the mean number of different adjectives on Days 6 to 9; second, show an increase in the mean number of different action verbs on Days 10 to 13; and third, have a mean number of different beginnings on Days 14 to 17 greater than other conditions, as well as having a mean number of different adjectives and a mean number of different action verbs that are less than the mean number when only these variables were under the contingencies, but are greater than baseline.

Of the six remaining students, the results of five students violate only one part of the definition of highly similar. The violation occurs either in the different adjective condition, when there is an increase in the mean number of action verbs as well as different adjectives for two of the students, or during the last condition when either the mean number of different adjectives or the mean number of different action verbs is not greater than the mean number during baseline, as for three of the five students.

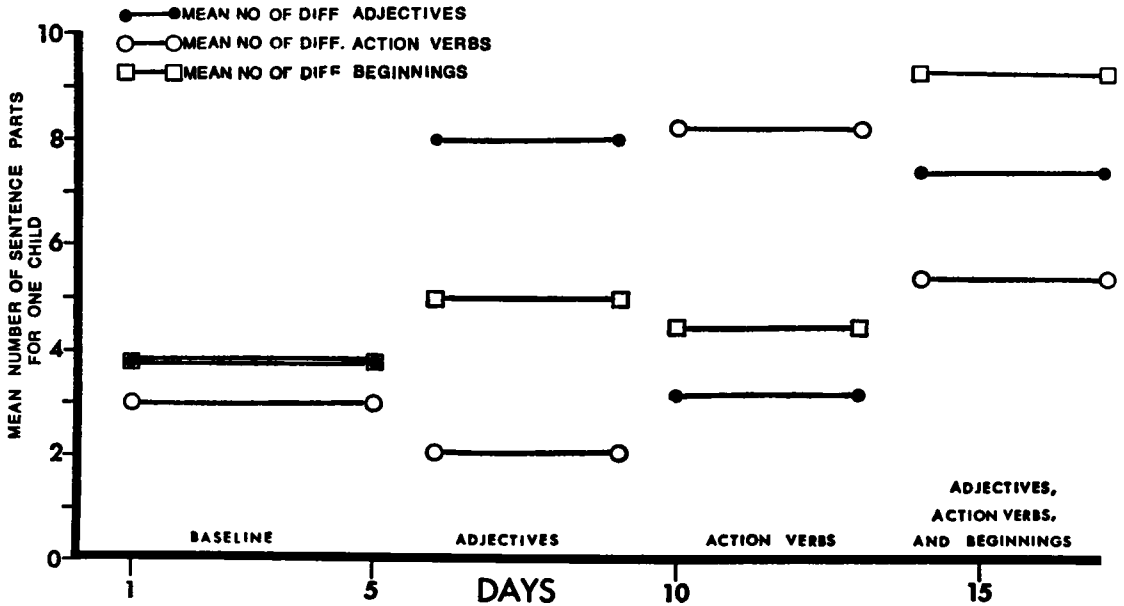


Fig. 3. Mean number of different adjectives, different action verbs, and different sentence beginnings for one child's stories for Days 1 to 17.

For one student, the mean number of different adjectives and different beginnings remained constant throughout the study.

Table 1 displays the combined results of the two independent raters, whose reliability was 46%. For the compositions written during base-

Table 1

Subjective ratings for creativity for one story from each condition for each child present during each condition, by two independent raters on an ordered scale of 1 to 4, from most creative to least creative.

	Most Creative	Second-Most Creative	Third-Most Creative	Least Creative
Baseline Condition	1	4	10	13
Different Adjective Condition	6	8	9	5
Different Action Verb Condition	16	5	4	3
Different Adjective, Action Verb, and Beginnings Condition	5	11	5	3

line, there were 13 ratings of least creative and only one rating of most creative. For compositions written during the different adjective condition, the majority of ratings were twos and threes. There were 16 ratings of most creative and only three ratings of least creative, when stories written during the different action verb condition were ranked. When compositions from the last condition were rated, the preponderance of ratings, 11, were in the category of second most creative. Two children were absent during the last condition, therefore the number of ratings totals only 24 (12 per rater) as opposed to 28 (14 per rater) in all other conditions in Table 1.

DISCUSSION

It is possible to modify qualitatively and quantitatively sentence structure and the usage of particular parts of speech by the use of prescribed procedures. Whether or not the changes in the objective compositional variables and sentence parts are unique to this set of procedures can not be determined from this study. Furthermore, some of the procedural aspects

that might warrant further research are the functions of giving examples before each session, of writing a statement of the contingency on the board, of varying the point values as well as the functions of team competition, of feedback in the form of points written on the board, and of the back-up of contingent points with extra recess and candy. The present research did not allow for the comparative evaluations of these different aspects of the procedures.

In this study, contingencies that required the use of different adjectives, different action verbs, and different sentence beginnings in 10-sentence stories, were more highly correlated with subjective judgements of creativity than were stories written during baseline, when no such contingencies were in effect. A more detailed inspection indicates that those stories written when there were contingencies requiring the use of different action verbs, were in general rated as more creative than those written when any other set of contingencies was in effect. However, since these results are from only two raters, and since rater results may also be specific to these procedures and stories, further systematic replication of the correlation is necessary.

The data indicate the existence of a possible response class with respect to usage of action verbs and adverbs. That is, even though contingencies were never directly placed on adverb usage, the number of adverbs written increased over baseline rates, and over the rates during the adjective condition when action verb usage was increased on Days 10 to 17. However, the use of a multiple baseline design indicates that there were no systematic changes in the frequency of occurrence of a number of other compositional responses, such as the mean number of compound sentences, prepositional phrases, and sentences with more than eight words, which were not related to the contingencies. This suggests that if practice effects were operating for the subject, they at least did not produce changes in all measures. Similarly, if order effects had

occurred and had influenced ratings, then the stories written during the last condition should have resulted in the greatest number of "most creative" ratings.

It should be emphasized that raters were instructed to employ only intra-subject judgements. Each student's story was compared only to other stories written by that same student. Therefore, no interpretation should be made that any of the stories are creative in comparison to absolute standards, or normative data on the set of all stories ever written by either elementary school children, or writers who are considered to be creative. However, it is possible that any productive and realistic attempt to develop creative writing skills of school children must deal with intra-subject approximations to creative productions, or most children will have little likelihood of writing absolutely or normatively creative stories.

Further research will be required to determine those specific contingencies and specific compositional variables that increase the probability that one story will be subjectively valued as more creative than another. It would also seem expedient to attempt to define operationally other response properties such as, "unusualness", "appropriateness", "transformation of the constraints of reality", and "condensation", purported to be essential for designating a behavior as creative (Jackson and Messick, 1965).

The results of the present study are in general agreement with those of Goetz and Baer (1971) and Goetz and Salmonson (1972), which indicate that variety of responses can be increased. Moreover, it is clear that the grammatical composition of children's written stories can be altered by relatively simple classroom procedures, and that the produced changes in composition are related to reliable, subjective judgements of creativity.

## REFERENCES

- Barrish, H. H., Saunders, M., and Wolf, M. M.  
 Good behavior game: effects of individual contingencies for group consequences on disruptive



- behavior in the classroom. *Journal of Applied Behavior Analysis*, 1969, 2, 119-124.
- Flescher, I. Anxiety and achievement of intellectually gifted and creatively gifted children. *Journal of Psychology*, 1963, 56, 251-268.
- Getzels, J. W. and Jackson, P. W. Family environment and cognitive style: a study of the sources of highly intelligent and highly creative adolescents. *American Sociological Review*, 1961, 26, (3), 351-359.
- Goetz, E. M. and Baer, D. M. Descriptive social reinforcement of "creative" blockbuilding by young children. In E. Ramp and B. L. Hopkins (Eds.), *A new direction for education: behavior analysis 1971*, Lawrence, Kansas: Support and Development Center for Follow Through, Department of Human Development, University of Kansas, 1971. Pp. 72-79.
- Goetz, E. M. and Salmonson, M. The effects of general and descriptive reinforcement on "creativity" in easel painting. In G. Semb (Ed.), *Behavior analysis and education—1972*. Lawrence, Kansas: Support and Development Center for Follow Through, Department of Human Development, University of Kansas, 1972. Pp. 53-61.
- Gutman, H. The biological roots of creativity. *Genetic Psychology Monographs*, 1961, 64, 419-458.
- Jackson, P. W. and Messick, S. The person, the product and the response: conceptual problems in the assessment of creativity. In J. Kagan (Ed.), *Creativity and learning*. Boston: Beacon Press, 1970. Pp. 20-35.
- MacKinnon, D. W. The study of creative persons: a method and some results. In J. Kagan (Ed.), *Creativity and learning*. Boston: Beacon Press, 1970. Pp. 20-35.
- Maltzman, I. On the training of originality. *Psychological Review*, 1960, 67, 229-242.
- Maslow, A. H. The creative attitude. In R. L. Mooney and T. A. Razik (Eds.), *Explorations in creativity*. New York: Harper and Row, 1967.
- Mooney, R. L. and Razik, T. A. (Eds.) *Explorations in creativity*. New York: Harper and Row, 1967.
- Osburn, A. F. *Applied imagination*. New York: Scribner's, 1957.
- Pryor, K. W., Haag, R., and O'Reilly, J. The creative porpoise: training for novel behavior. *Journal of the Experimental Analysis of Behavior*, 1969, 12, 4, 653-662.
- Torrance, E. P. and Tan, C. A. *Revised scoring guide for analyzing attitudes concerning divergency in imagination stories*. University of Minnesota, 1964.

Received 13 June 1972.

(Revision requested 18 September 1972.)

(Final acceptance 7 May 1973.)