8-1982

The Development of an Expressive Language Scoring System

Pamela Tamme
Western Kentucky University

Follow this and additional works at: https://digitalcommons.wku.edu/theses
Part of the Child Psychology Commons, and the School Psychology Commons

Recommended Citation
https://digitalcommons.wku.edu/theses/2892

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Masters' Theses & Specialist Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.
Tamme,

Pamela Steagall

1982
THE DEVELOPMENT OF AN EXPRESSIVE LANGUAGE SCORING SYSTEM

A Thesis
Presented to the Faculty
Department of Psychology
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts

by
Pamela Steagall Tamme
August 1982
AUTHORIZATION FOR USE OF THESIS

Permission is hereby

☑ granted to the Western Kentucky University Library to
make, or allow to be made photocopies, microfilm or other
copies of this thesis for appropriate research or scholarly
purposes.

☐ reserved to the author for the making of any copies of this
thesis except for brief sections for research or scholarly
purposes.

Signed  

Pamela Skagall Farnone

Date  

July 14, 1982

Please place an "X" in the appropriate box.

This form will be filed with the original of the thesis and will control
future use of the thesis.
THE DEVELOPMENT OF AN EXPRESSIVE
LANGUAGE SCORING SYSTEM

Recommended July 12, 1982

[Signatures]

[Signatures]

Approved 7-20-82

[Signature]

Dean of the Graduate College
# TABLE OF CONTENTS

| LIST OF TABLES | Page  
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF APPENDIXES</td>
<td>vi</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vii</td>
</tr>
</tbody>
</table>

**Chapter**

I. INTRODUCTION  
II. REVIEW OF THE LITERATURE  
Elaboration  
Elaboration as a Measure of Creativity  
Elaboration and Depth of Processing  
Summary  
III. METHOD  
Subjects  
Selection of Students and Task Assignment  
Selection of Teachers and Task Assignment  
Procedure  
Development of Scoring System  
Training of Scorers and Reliability Study  
Analysis  
IV. RESULTS  
Inter-judge Reliability  
Intra-judge Reliability  
Story-starter Content and Context  
General Linear Regression Analyses  
Stepwise Procedures  
V. DISCUSSION  
Inter-judge Reliability  
Intra-judge Reliability  
Story-starter Content and Context  
General Linear Regression Analyses  
Stepwise Procedures  
Implications  

iii
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Chi-square with Content as the Independent Variable and Teacher Ratings of Story Creativity as the Dependent Variable</td>
<td>28</td>
</tr>
<tr>
<td>2.</td>
<td>Chi-square with Content as the Independent Variable and Judges' Scores as the Dependent Variable</td>
<td>29</td>
</tr>
<tr>
<td>3.</td>
<td>Chi-square with Context as the Independent Variable and Teacher Ratings of Story Creativity as the Dependent Variable</td>
<td>30</td>
</tr>
<tr>
<td>4.</td>
<td>Chi-square with Context as the Independent Variable and Judges' Scores as the Dependent Variable</td>
<td>32</td>
</tr>
<tr>
<td>5.</td>
<td>General Linear Regression with Teacher Ratings of Story Creativity as the Criterion Variable</td>
<td>33</td>
</tr>
<tr>
<td>6.</td>
<td>General Linear Regression with Teacher Ratings of Student Creativity as the Criterion Variable</td>
<td>35</td>
</tr>
<tr>
<td>7.</td>
<td>Stepwise Procedure using Teacher Ratings of Story Creativity as the Criterion Variable</td>
<td>36</td>
</tr>
<tr>
<td>8.</td>
<td>Stepwise Procedure using Teacher Ratings of Student Creativity as the Criterion Variable</td>
<td>37</td>
</tr>
<tr>
<td>Appendix</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>A. Story-starter Matrix</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>B. Criteria for Teacher Ratings of Stories</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>C. Teacher-generated Criteria for Creativity</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>D. Directions for using the Expressive Language Scoring System</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>E. Intra-judge Reliability Table</td>
<td>61</td>
<td></td>
</tr>
</tbody>
</table>
This study was designed to develop an expressive language scoring system so that the creative quality of a person's written language could be quantified and measured. Sixty fifth-grade students, 31 males and 29 females, were given the opening line, i.e., story-starter, to a story and were asked to finish the stories. Each student received two story-starters concerning the same object ("content")—either box, string or money—but were different in "context" (usual or unusual setting). Thus, a total of 120 stories were written, with 40 stories being from each content group. These stories were then rated according to their level of creativity by ten teachers having fourth, fifth or sixth grade teaching experience. These teachers were unfamiliar with the students who wrote the stories. The teachers were divided into two groups of five, and each group rated 60 stories on a seven point scale. The stories were divided equally so that each group rated one story from each child and so that the content and context of the stories were balanced. After the teachers rated the stories, they were asked to list the criteria they used for rating the stories. In addition to the teacher ratings of creativity, each student who wrote a pair of stories was rated on general creativity by his/her classroom teacher. The teachers' lists of criteria, along with past research by Guilford (1968) and Torrance (1974), helped determine what to include in the present exper-
imenter-developed scoring system. There were seven sub-factors used, viz., ideational fluency, associational fluency, elaboration, relevant flexibility, irrelevant flexibility, originality, and organization. The subscale scores were combined to yield a total score for each story. The relationship among the total scores, sub-factor scores, teacher ratings of story creativity and classroom teacher ratings of student creativity was analyzed using regression analyses. Results showed that the total score was the best predictor for teacher ratings. The correlation between total score and teacher ratings of story creativity was .67. This suggests that the experimenter-developed scoring system validly reflects teacher ratings of story creativity. The correlation between the scoring system and classroom teacher ratings of student creativity was .36. The low correlation between the scoring system and classroom teacher ratings could be due to biases from knowing the students and other factors which could have influenced their judgements. Another possibility is the indication that teachers are, in general, poor judges of creativity.
CHAPTER I

INTRODUCTION

There are various ways to measure achievement, such as standardized tests and cumulative grade point averages. However, perhaps the most common way that achievement is judged is informally via teacher judgement of a student based upon his/her classroom performance. Performance within the school environment is manifested in a number of ways as determined by teacher assignments, e.g., math computations and oral and written responses. A majority of classroom tasks (e.g., spelling tests, dictation, short answer questions, essay questions) require verbal responses, and most often these responses are in written form (Dunkin and Biddle, 1974). In many instances, however, students have difficulty expressing themselves in writing. For example, a child may not be able to answer an essay question or write a satisfying creative story because he/she cannot formulate important ideas necessary for an adequate response (Ebel, 1979).

Hence, achievement within the school environment may be highly dependent upon a student's ability to express himself/herself.

In this study, the manner in which a person expresses himself/herself is considered to be partially dependent on his/her ability to elaborate upon ideas. This perspective differs from traditional investigations of elaboration (Stein, Morris and Bransford, 1978). Elaboration has been viewed from two different, yet complimentary,
approaches--its role in input, and its role in output.

Ideally, by studying either input information or the resultant output, one can indirectly measure the cognitive processes which occur somewhere between placing information into memory, i.e., input, and recalling the information from memory, i.e., output. Bransford (1979) defines elaboration as the quality and quantity of semantically congruent and relevant information that follows a base phrase or sentence. This view of elaboration stresses the importance of elaborative input on output, i.e., recall. As implied in this definition, Bransford believes that not only the amount of elaboration during input, but also the type, e.g., precise/imprecise or effective/ineffective, affects recall. For example, "The diamond was too expensive for the poor man" is effectively elaborated because poor relates to expensive. "The diamond was too expensive for the slow man" is not effectively elaborated because slow does not relate to expensive. Thus, recall is facilitated more easily by the former sentence.

While Bransford has emphasized the elaborative aspects of input to recall of textual information, others have emphasized the elaborative aspects of output (Guilford, 1968; Torrance, 1970). Elaboration, according to Guilford and Torrance, is only one factor which contributes to the total quality of a person's output. They define elaboration as the ability to take what is already a fairly well-rounded product and expand it with details. Teacher judgement of classroom performance is based largely on student output. Thus, the present study intends to carefully examine student output rather than teacher or text input. Since one of the most common modes of
student output is expository written language, it is important to examine the qualities of that language. Torrance's notion of elaboration and three other variables—fluency, flexibility and originality—based on Guilford's (1968) model of intellect, have some implications for examining the quality of student-generated expressive language. These variables, according to Torrance, are measures of creative thinking which has been expressed in written form and involve both the quality and the quantity of written language. Torrance has devised scoring norms for these variables which makes it easier to examine and measure a student's written expressive language. One major drawback to Torrance's method of scoring creative language is that he asked subjects to list their responses to his tasks within his Tests of Creative Thinking (1974). This did not allow for a "flow" of language which can be found by having a subject write a complete paragraph or story.

In the present study, elaboration is broadly defined as the quality of expressive language. This definition is not to be confused with the narrower definitions provided by Bransford (1979), Guilford (1968), and Torrance (1970). Bransford's definition focuses on the effects of elaborative input on recall, particularly from text. Elaboration as defined by Guilford and Torrance constitutes but one factor contributing to the present definition. The present study views the definition of elaboration provided by Guilford and Torrance along with fluency, flexibility and originality, which are also defined by Guilford and Torrance, and an experimenter-defined factor, organization, as contributing to the overall quality of language.
In order to more clearly understand the relationship between elaboration and achievement, it is necessary to quantify and analyze the factors contributing to the quality of expressive language. Research has demonstrated that it is not only the quantity, but also the quality, of elaborative input that affects performance (Stein, Morris and Bransford, 1978). However, the quality of elaborative output, as defined here, has not been quantified so that the relationship between achievement and the quality of elaborative output can be determined. There are models (Kintsch, 1974) which analyze the conceptual meaning of language, which is a step in this direction. These models, however, are based on direct recall from text, which is not the same as recalling information and using it in new forms or new connections, i.e., transfer recall (Guilford, 1968). Expressive language within the classroom is seldom, if ever, word-for-word text recall. Instead, it is a more open-ended process which allows for transfer recall to occur.

The purpose of this study was to develop a system for quantifying elaboration, which is broadly defined as expressive language. Once this system has been developed, further research in the area of using expressive language as a measure of elaboration can take place. For example, the relationship between elaboration and classroom achievement and the differences between oral and written elaboration can be studied. These future studies, along with the present one, may help teachers within the school environment increase their awareness of the importance of effective elaborative abilities.
CHAPTER II

REVIEW OF THE LITERATURE

A person can have conscious control of his/her own cognitive actions which involve such processes as predicting, checking, monitoring, reality testing, coordinating and controlling deliberate attempts to study, learn or solve problems (Brown, 1980). Monitoring one's own state of knowledge occurs only after a person develops and begins to realize that he/she is an active agent in knowing. Therefore, young children are more limited in their knowledge about their own cognitive abilities than are adults. That is, they are not as aware of their ability to monitor their own state of knowledge. According to Brown, as a child grows and experiences education, he/she has the opportunity to become more aware of the ability to control his/her cognitive actions and the importance of this in learning. It would seem that the ability to elaborate, which involves memory, comprehension and language acquisition, is also enhanced when a person becomes aware of his/her own abilities in this particular area of learning. Thus, within the school environment it would be to a student's advantage to become aware of his/her ability to learn various skills, such as elaborative processes, which lead to achievement and academic success.

Elaboration

Elaboration has been studied from two basic perspectives: (1)
its role as an expressive output measure for creativity (Guilford, 1968; Torrance, 1974) and (2) its input role as a memory facilitator (Craik and Lockhart, 1972; Stein, Morris and Bransford, 1978). Elaboration historically has been one factor used in the measurement of creativity. When measuring creativity through output, the subject has freedom to respond as he/she chooses, and the elaboration generated is not controlled by the experimenter. However, when looking at elaboration as a facilitator of memory, the experimenter controls a subject's ability to recall information from memory. Both of these perspectives, the role elaboration has as an expressive output measure for creativity and as a memory facilitator, are of value to the present study--the former because of its open-ended method of measuring elaboration and the latter because of its view of elaboration as a way to assess what a subject has retained in memory.

Elaboration as a Measure of Creativity

One of the primary reasons that the creative output of elaboration is important is because it is not as limited as the memory/input aspect of elaboration. For example, creative output of elaboration allows a subject the freedom to respond at any length he/she chooses and in any way that he/she wants. Secondly, creativity is often studied by creative writing and expressive language. Elaboration, for purposes of this study, will also be studied using expressive language. Finally, elaboration from this point of view is an output, i.e., a result.

Guilford (1968, 1971, 1973, 1975) studied the role that creativity plays in the structure-of-intellect (SI). In order to better understand how creativity fits into Guilford's model, it is necessary
to briefly explain his SI model. The SI model is a complex, thorough one in which creativity plays a significant part, and creativity, according to Guilford, involves a variety of mental functions. The SI model is best illustrated by a cube-like form where height, length and width represent three distinct abilities of mental functioning. These three abilities are operation, content and product.

**Operation.** Mental abilities can be classified by their type of operation, or the basic kind of process which is performed. There are five intellectual operations, viz., cognition, memory, divergent production, convergent production, evaluation. Of these five operations, divergent production and convergent production are of particular importance because classroom tasks often demand a student to recall information in these two forms. Both divergent production and convergent production depend on retrieval of information from storage (Guilford, 1975). However, tasks involving divergent production operations are ones where a number of different, alternative productions are possible. There can be more than one "right" answer to a question or response. In convergent production, the information given is restricted in such a way that there is only one possible answer or response. It is important to remember that divergent production abilities are the most relevant to creative thinking, but creativity is not limited to only tasks requiring divergent production abilities (Guilford, 1968).

**Content.** The SI model also specifies five kinds of content, viz., visual-figural, auditory-figural, symbolic, semantic and behavioral. The five kinds of content represent the separation of the brain functions and the brain structures used for different types
of intelligence, e.g., visual, auditory, semantic. In other words, different areas of the brain are responsible for carrying out different functions that a person does. The most important type of content in the present study is semantic, e.g., sentences, because the data will be collected in sentence form. Any information which is processed by thoughts rather than signs, perceptions or visual/auditory modes is semantic in content.

**Product.** The third, and last, main component found within Guilford's SI cube is the structure of information, or product. The importance of the product in the SI model is that it allows a person to discriminate in learning. There are six products, viz., unit, class, relation, system, transformation, implication. The system is defined as an original group of wholes which are related to one another, e.g., a sentence. Because the present study will use sentences as the mode of expression, the system is the most relevant product.

Within Guilford's (1968) model, creativity extends beyond the domain of intelligence. Thus, an IQ test does not necessarily allow for an adequate measure of creativity. Guilford has found that there are four basic factors found within creativity. These factors are fluency, flexibility, originality and elaboration. The above factors belong in the divergent thinking category because the subject has to generate a variety of answers. General definitions of each of the factors are as follows.

**Fluency.** Guilford (1968) has found three fluency factors within his SI model. These factors are ideational fluency, which is the rate of generation of a quantity of ideas; associational fluency, which
pertains to the completion of relationships; and expressional fluency, which pertains to the facile construction of sentences.

**Flexibility.** Two abilities for flexibility, spontaneous flexibility and adaptive flexibility, have been recognized by Guilford. Spontaneous flexibility has to do with changes in direction of thinking when one has not been instructed or does not need to change directions. Adaptive flexibility involves changes in direction of thinking in order to solve problems (Guilford, 1975).

**Originality.** Originality, according to Guilford, involves semantic transformations, i.e., reinterpreting, revising or reorganizing something, such as a story. A transformation is defined as a change of some sort, and transformations are placed into memory so that they can be used again.

**Elaboration.** The final factor, elaboration, is thought to be important and relevant in planning. Elaboration is defined as the ability to take what is already a fairly well-rounded product and expand it with details. Thus, a subject who can elaborate well is able to produce detailed steps for a particular project or situation.

Generally low correlations have been found between traditional IQ tests and tests of creative potential. For example, in a study by Guilford and Hoepfner (1966), there was a correlation of .32 between divergent production tasks and IQ. There are few persons who have above average creative abilities and below average IQs, but there are many cases of persons with lower than average creative abilities and higher than average IQs (Guilford, 1975). This brings up the question as to whether or not the school environment adequately allows for development of creative abilities, or is the school environ-
ment using different criteria when defining and assessing creative abilities?

Creativity, in general, is a form of problem-solving, but it is not yet known whether all forms of creative thinking involve problem-solving (Guilford, 1968). One type of creative thinking is known to be a problem-solving activity, and that is creative writing. Torrance (1962) has found that writing is one way to measure creative thinking. Several of his ideas and one of his Tests of Creative Thinking (1974) are variations of Guilford's work.

The Torrance Tests of Creative Thinking include four batteries of test activities, two verbal and two figural. Each of Torrance's tests yield three or four scores: a fluency score, a flexibility score, an originality score, and--for figural tests only--an elaboration score. Torrance has measured verbal elaboration but did not use elaboration when norming the verbal ability scores. Of the verbal and figural tests, the verbal tests are more relevant to the present study. This is because of the heavy influence on verbal abilities in the classroom. This is not to say that figural abilities are not used in the classroom, but the present study is restricting its focus to only the verbal abilities.

Torrance's verbal fluency score reflects a person's ability to produce a large number of ideas with words. The verbal flexibility score reflects a person's ability to produce a variety of kinds of ideas, to shift from one approach to another, or to use different types of strategies. Torrance defines the verbal originality score as a person's ability to produce ideas that are away from the obvious or commonplace. Though he did not establish scoring norms
for verbal elaboration, much of Torrance's view of the factor is based heavily on Guilford's (1968) previous work with elaboration.

According to Guilford, elaboration occurs in divergent thinking sequences where individuals are able to freely generate their own ideas. Guilford describes elaboration as being either structured or free and states that elaboration occurs when an individual builds upon a point already made or develops a point but does not shift to a new point (Dunkin and Biddle, 1974).

Elaboration is encouraged by the directions to Torrance's tests. For example, directions for the figural tests urge the students to add ideas so that the pictures will tell as complete and as interesting a story as possible. Elaboration is encouraged in the directions of the verbal tasks in much the same way. High scores in verbal elaboration tasks seem to indicate school achievement (Torrance, 1974). Low scores on these same tasks appear to be characteristic of "underachievers" in school. Hence, school achievement, according to Torrance, is related to verbal elaboration.

Torrance (1974) also found that some students produce a large number of ideas but do not elaborate any of them very well, whereas other students produce very few ideas but make them very elaborate or fancy. This suggests the quality versus quantity aspect of elaboration as defined by Bransford (1980), which will be discussed in greater length in the following section on elaboration and depth of processing.

Finally, Torrance (1962) has found that students who are creative but have relatively low IQs achieve as well in school as students who are higher in intelligence by 20 points. These findings
suggest that an IQ score may not always be as good a predictor of classroom achievement as tests of creativity, such as creative verbal tasks.

Looking at the creative aspects of elaboration, when elaboration is defined as the quality of expressive language, shows how important it is for students to be able to elaborate in the classroom. School performance is judged more heavily on verbal abilities than it is on nonverbal abilities (Ebel, 1979). Therefore, it would seem that students who can effectively elaborate are the ones who achieve and succeed in the school environment.

Elaboration and Depth of Processing

The term elaboration is often used when studying how information is processed into memory, i.e., input, and how information is recalled from memory, i.e., output. It is important to look at the role that elaboration plays in memory because, though it may appear different from creative elaboration, it is essentially the same.

One theoretical framework proposed by Craik and Lockhart (1972), the "Depth of Processing" theory, considers elaboration as one way to enhance a subject's memory and ability to recall information. Depth of processing is a hierarchy of cognitive processing stages which facilitate storage and retrieval of information. "Depth" refers to degrees of semantic or cognitive analysis (Craik and Lockhart, 1972). Craik and Lockhart suggest that memory is a function of depth of processing, e.g., deeper or higher levels of processing are associated with more elaborate, longer lasting, and stronger memories. Shallower or lower states of processing involve analyzing physical or sensory features, e.g., loudness, brightness, lines, or
stimuli such as a word, whereas higher levels of processing involve pattern recognition and meaning of stimuli. Likewise, the higher, or better, that information is stored in a subject's memory, the more a subject can later recall. Elaboration is considered a high level of processing information and occurs after a stimulus has been recognized. The stimulus, e.g., a word, sets off further associations, images or stories which are based on a subject's past experience with the stimulus. Therefore, elaborative output is based upon the degree to which a subject has placed information into memory storage. Stimuli which are meaningful and familiar to a subject are processed into a deep level more readily than less meaningful stimuli and thus will be better retained and recalled.

The process of elaboration has been studied by Stein, Morris and Bransford (1978). They found that subjects who encode, or place into memory, information in a more "superficial", nonsemantic way sometimes have greater retention than subjects who encode information using "deeper", semantic encodings. This indicates that in certain testing situations, nonsemantic tasks can be superior to semantic tasks. Reasons for this could be due not only to the quality of information stored by a subject but also to the prior knowledge of a subject and the ability of a test to adequately assess the information that has been stored by the subject. The findings of Stein et al (1978) study imply that analyzing the ability to recall information from memory based on semantic versus nonsemantic processing is not as important as looking at how individuals encode and use what they know, be it information that is encoded semantically or non-semantic.
Reder (1980) has done an extensive study concerning the role that elaboration plays in comprehension and retention. Elaboration, according to Reder, occurs when a person does extra processing of data which leads to additional, related or redundant propositions that facilitate better memory of data. The two main methods used to measure both comprehension and retention in the study were question answering and free recall. Various methods of analyzing the recalled responses were discussed. The script-elaboration model discussed in Reder's paper (1980) views elaboration as a way of encoding prose so as to facilitate memory. In the script-elaboration model, the process of elaboration aids in long-term retention of the input, i.e., if the input is well elaborated during reading then recall is facilitated. Past experiences of a student, comprehension of the text, interest in the subject matter, concentration, time allowed for reading and inherent tendency for a student to elaborate are all factors which influence memory.

Each of the above mentioned studies deals with elaboration as a facilitator of memory and, thus, they primarily use recall as a way to measure how well a person has stored information. Emphasis is placed on using effective elaboration when placing information into memory, e.g., input, so that later the information can be better recalled from storage. Stein, Morris and Bransford (1978) and Craik and Lockhart (1972) have studied the recall of single words, whereas Reder (1980) has reviewed various studies of prose recall. Each of these studies consider recall as being influenced by how well a subject has initially encoded information. Recall which has been elaborated upon is thought to have been better stored into memory when
initially encoded.

Basically, elaboration works in two ways which compliment each other: (1) information that is effectively elaborated to a subject is better remembered than information that is not effectively elaborated to him/her and (2) information that is encoded more effectively is likely to be information upon which a subject can elaborate later when recalling the information. Therefore, if a teacher can effectively elaborate when teaching, subjects will be able to more easily store the information into memory. Likewise, when asked to use more elaborative output upon the recalled information.

The need for effective elaborative input in the classroom can clearly be seen by the classroom teacher-student relationship. A teacher must be able to effectively teach so that the information which is stored by students can be recalled when demanded by tasks requiring memory or transfer recall (i.e., using recalled information in new ways). Tasks requiring recall of a single stimulus, such as a word (e.g., fill-in-the-blank), and tasks requiring recall of prose (e.g., essay questions) are often used in the classroom (Brown, Campione and Day, 1981). Better memory for the information processed logically leads to better achievement. Hence, it is felt that elaboration, as a facilitator of memory, affects classroom achievement, and what is expressed through output (e.g., essay questions, creative stories) is thought to be a reflection of how well information was initially stored in memory.

Though both direct recall and transfer recall are important in learning, it is felt, in the present study, that the latter is more important. Transfer recall allows a person to take direct recall
and use it in new forms or new connections, which gives opportunity for more creative elaboration as defined by Guilford (1968) and Torrance (1974).

Summary

In summary, the above literature review gives a broad overview of various areas of research which are, at least in part, important to the present study. In the present study elaboration will be measured through analyzing written expressive language. Torrance (1974) used written expressive language as a way to measure creative thinking, but he was somewhat limited in his scope because he had subjects list only responses, which breaks the flow in writing. His results did show, however, that verbal elaboration, as he defined it, is positively related to classroom achievement. It is important to remember, however, that Torrance's definition of elaboration is only a small part of the present study's definition. Craik and Lockhart (1972) believe elaboration occurs as a result of information being placed higher into memory storage. Thus, a person who elaborates well is one that has stored the information well. As Stein et al. (1978) demonstrated, the manner in which information is stored into memory is more important than what type of information is stored. There are several factors which can enhance or hinder the storage of information, and one such factor is the classroom teacher.

Many linguistic modes are used in the classroom environment (Dunkin and Biddle, 1974). Language is the primary way in which teachers instruct their students, and language--both oral and written--is the way students primarily communicate what they have learned.
The present study will attempt to establish a means by which the quality of expressive language, i.e., elaboration, can be quantified. Factors similar to those defined by Guilford (1968) and Torrance (1974) as well as teacher-generated criteria for creativity were used to develop the present expressive language scoring system. Once the quantification system is developed, further research concerning elaboration and its role in school achievement can be explored. For example, is elaboration a better predictor of school achievement than IQ scores, and are students who are gifted according to their IQ score but are underachieving in the classroom poor elaborators?
CHAPTER III

METHOD

Subjects

Selection of Subjects and Task Assignment

Three fifth-grade classrooms were selected from an available pool of five fifth-grade classrooms in two elementary schools in south central Kentucky. The sample consisted of 31 males and 29 females (N=60). Each student in the sample was asked to write two stories. Each student was given an opening line for each story, i.e., a story-starter. The story-starters were organized into one of six categories: (1) a story about a box in a usual situation, (2) a story about a box in an unusual situation, (3) a story about a string in a usual situation, (4) a story about a string in an unusual situation, (5) a story about money in a usual situation, and (6) a story about money in an unusual situation. The stories were presented in alternating order to the students in order to control for context of the stories, i.e., usual or unusual. The story-starters appear in Appendix A.

The students were instructed to read the story-starters and then to help complete the story. Of the sixty students, one-third responded to the box story-starter in usual and unusual context (N=20), one-third responded to the money story-starter in usual and unusual context (N=20), and one-third responded to the string story-
starter in usual and unusual context (N=20). Thus, a total of 120 stories were collected. The students were allowed 10 minutes in which to write each story based upon the story-starters (a total of 20 minutes per student).

Selection of Teachers and Task Assignment

Rating of stories. Ten teachers who were either currently teaching fourth, fifth or sixth grade or had taught one of these grades in the past were paid ten dollars each to help with the present study. These teachers came from one of two night classes taught at Western Kentucky University and were selected on a first-come first-serve basis. Each teacher was asked to rate 60 stories, one from each student, using a seven point scale based on the creativity level of each story. The directions for this rating system can be found in Appendix B. Five teachers rated the same 60 stories, and the other five teachers rated the remaining 60 stories. This was done in order to obtain five creativity scores for each story. The five scores could then be averaged in order to yield one overall score for each story. Each teacher received an equal number of stories within each content by context area (i.e., each teacher received ten stories from each of the six categories shown in Appendix A). The teachers were then asked to list the criteria they used for rating the stories.

Rating of students. In addition to the ten teacher ratings of story creativity, each student who wrote a pair of stories was rated on his/her level of general creativity by the classroom teacher. These teachers did not read the stories but based their ratings on general classroom interaction with the students.
Development of the Scoring System

A frequency count was obtained of the listed criteria that the teachers used to discriminate between the seven story categories. This was done in order to assess the degree of agreement between the teachers' listed criteria of creativity and the variables found within Torrance's scoring system. As can be seen in Appendix C, the teachers' listed criteria closely resembled the variables that Torrance used in his scoring system.

The present study's scoring system was based on five factors, all designated as important by the teachers' listed criteria. Four of the factors were verified with Torrance (1974) and Guilford (1968) as being important aspects of creativity (viz., fluency, flexibility, originality, elaboration). One factor in the present scoring system, organization, got at Guilford's notion of system building. Two of the factors, fluency and flexibility, were divided into sub-categories. Fluency was divided into ideational and associational fluency, and flexibility was divided into relevant and irrelevant flexibility. Thus, there were seven sub-scores which, when added together, yielded a total score for each story.

Operational definitions of each factor were created by the experimenter so that quantification scores for each story could be obtained. The definitions are as follows.

Ideational fluency referred to the use of an object as measured by the number of purposes served by the object. Each purpose could be counted only once within a story. A subject's ideational fluency score, then, was determined by counting the total number of purposes
served by the object within each story.

Associational fluency was defined as the number of things done with, to, or by each object and/or the consequence of the usage. This could be determined by asking the questions: What has been done with the object? What are the uses to which it is put? What was done to it if it was transformed, i.e., changed? A subject's associational fluency score equaled his/her total number based on the above definition of associational fluency.

Elaboration was measured by counting all of the adjectives, adverbs, prepositional phrases and other descriptors and qualifiers that were not needed to complete a thought, sentence or basic idea. A subject's elaboration score equaled the total number of descriptors and qualifiers found within his/her story.

Relevant flexibility was defined as the number of basic ideas or subthemes contained within the story which were consistent with the overall theme of the story. Subtheme changes were indicated by changes in action, perception or thinking on part of the author or story character.

Irrelevant flexibility is similar to relevant flexibility except that it was defined as the number of basic ideas or subthemes contained within the story which were inconsistent with the overall theme of the story. Subtheme changes were indicated by changes in action, perception or thinking on part of the author or story character. A subject's irrelevant flexibility score equaled the total number of irrelevant basic ideas or subthemes within his/her story which is then subtracted from the subject's overall score.

Originality was defined as the uniqueness of the main, overall
usage of the object. The full story was read in order to determine
the primary usage or theme. The stories were divided into their six
categories and the primary usage was determined for each within its
group. Responses, i.e., usages, which were found in only one story
out of the possible 20 received a score of four. Responses found
within two stories received a score of three. Responses found within
three stories received a score of two. Responses found within four
stories received a score of one, and responses found within five or
more of the stories received a score of zero. If, however, the re-
sponse was a transformation, i.e., the object was taken and changed
from a typical usage, it received an additional bonus point. Also,
a bonus point was given for a creative twist, i.e., a surprise end-
ing, at the end of the story. Thus, a subject's originality score
could vary between zero and six.

Organization was defined in the following way. Beginning with
the story-starter, the number of sentences were counted which were re-
lated to the prior sentence. If a sentence was related to the prior
sentence, it received a point. If a sentence was not related to the
prior sentence, it did not receive a point. The relationship was
judged by asking the question "Is the idea of the sentence related
to the prior sentence?" The score equaled the total number of points
received for related sentences divided by the total number of sen-
tences then multiplied by 10. The story-starter did not count as
a sentence, but the first phrase or sentence written by the subject
was evaluated for relationship to the story-starter.

Finally, a subject's total score was equal to the sum of the
seven sub-factors when added together--with the exception of ir-
relevant flexibility, which is subtracted. A copy of the directions used by the scorers, along with scoring examples, can be found in Appendix D.

Training of Scorers and Reliability Study

A workshop was held in order to train five judges who were to use the above operational definitions of ideational and associational fluency, relevant and irrelevant flexibility, originality, elaboration and organization to score the stories. Each story had to be read and scored seven times—once for each variable. Next, a total score for each story was obtained. A practice session was held where the judges scored seven of the same stories together. After the stories were scored and it was felt that the scoring system was understood by each judge, the judges scored 30 of the same stories separately in order to check for inter-judge reliability. Intra-judge reliability was established by having each judge rescore 10 randomly selected stories from the original pool of 30 stories after a two month time lapse.

Analysis

In addition to the inter-judge and intra-judge reliability checks, it was necessary to determine whether or not content or context influenced the teacher ratings of story creativity and judges' scores. Therefore, a check was done using the chi-square statistic. It was hoped that neither teachers nor judges would place stories, based on the seven point scale, in categories according to the content or context of the stories more often than would be expected by chance.
The relationship among teacher ratings of story creativity, teacher ratings of student creativity and judges' scores was analyzed using linear regression analysis. The criterion variables were teacher ratings of story creativity and teacher ratings of student creativity. The predictor variables were the total score and sub-factor scores on the expressive language instrument. These analyses helped to determine if the experimenter-designed scoring system was measuring the same thing as the teacher raters and the classroom teachers when they judged the creativity level of students.
CHAPTER IV

RESULTS

One of the main goals of the present study was to establish the degree of inter-judge reliability and intra-judge reliability for the expressive language scoring system. The average correlation among the judges, based on the scores of 30 stories (i.e., inter-judge reliability) and a judge's percent of agreement with himself/herself over time (i.e., intra-judge reliability), was both necessary to obtain so that the consistency of scoring stories within and among the judges could be determined.

The chi-square statistic was used to determine if the teachers who rated the stories and the judges who scored the stories rated or scored them in such a way that more stories of a particular content or context received a particular score (i.e., consistently higher or lower) more often than would be expected by chance.

Two basic analyses were performed on the data using the SAS package ("The SAS Users," 1979). General linear regression and step-wise procedures were used to investigate the predictor variables (viz., total score, ideational fluency, associational fluency, elaboration, relevant flexibility, irrelevant flexibility, originality, organization) and each of the criterion variables (viz., teacher ratings of story creativity and teacher ratings of student creativity). The general linear regression analysis shows how much a particular predictor variable affects the criterion variable, given
all the other predictor variables. The stepwise procedure indicates the best combination of predictors. A total of four regression analyses were performed.

Inter-judge Reliability

After scoring seven stories in a practice session, five judges scored 30 of the same stories which were balanced for content and context. This was done in order to determine the degree of inter-judge reliability when using the expressive language scoring system. An average correlation of .80 was obtained among the judges. It was felt that a correlation of .80 was high enough to obtain reliable scores, and the experimenter scored the remaining 90 stories.

Intra-judge Reliability

In order to check for intra-judge reliability, each of the five trained judges rescored, after two months, 10 randomly selected stories from the original pool of 30 stories used to establish inter-judge reliability. Total score intra-judge reliability ranged from .89 to .93. The intra-judge reliability coefficients for each of the subscales appear in Appendix E.

Story-starter Content and Context

A total of four chi-square analyses were used to determine whether teachers and judges categorized a greater number of stories than would be expected by chance into each of the seven creativity rating categories. The first two chi-square analyses dealt with content of the story-starters (i.e., box, money or string). The last two chi-square analyses dealt with context of the story-starters.
(i.e., usual or unusual).

In the first analysis, the independent variable was content of the story-starter; the categorical dependent variable was teacher rating of stories (viz., very poor vs. poor vs. fair vs. average vs. good vs. very good vs. superior). It was found that content influenced teacher ratings of story creativity and was significant, $x^2(12)=31.75; p<.01$. It seemed that the stories about a box were rated higher on the seven point scale than stories about string or money as reflected by Table 1. Thus, teachers may have been influenced by the content of the stories.

In the second chi-square analysis, the independent variable was content of the story-starter; the categorical dependent variable was judges' scores of the stories. The judges' scores were placed in the same seven categories (viz., very poor vs. poor vs. fair vs. average vs. good vs. very good vs. superior) according to the range of the judges' scores. It was found that the judges did not place stories in any of the seven categories more often than would be expected by chance, $x^2(12)=16.059, p>.05$. The data are summarized in Table 2.

The third chi-square analysis dealt with context of the story-starters (i.e., usual vs. unusual). The independent variable was context of the story-starter; the categorical dependent variable was teacher ratings of stories (viz., very poor vs. poor vs. fair vs. average vs. good vs. very good vs. superior). It was found that the teachers did not place stories in any of the seven categories more often than would be expected by chance, $x^2(6)=12.09, p>.05$. See Table 3 for a summary of the data.
Table 1
Chi-square with Content as the Independent Variable
and Teacher Ratings of Story Creativity as the Dependent Variable

<table>
<thead>
<tr>
<th>Rating: Content</th>
<th>Very Poor</th>
<th>Poor</th>
<th>Fair</th>
<th>Average</th>
<th>Good</th>
<th>Very Good</th>
<th>Superior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box</td>
<td>30</td>
<td>16</td>
<td>26</td>
<td>34</td>
<td>32</td>
<td>36</td>
<td>26</td>
</tr>
<tr>
<td>String</td>
<td>46</td>
<td>25</td>
<td>26</td>
<td>31</td>
<td>32</td>
<td>23</td>
<td>17</td>
</tr>
<tr>
<td>Money</td>
<td>25</td>
<td>33</td>
<td>39</td>
<td>45</td>
<td>27</td>
<td>18</td>
<td>13</td>
</tr>
</tbody>
</table>
Table 2
Chi-square with Content as the Independent Variable and Judges' Scores as the Dependent Variable

<table>
<thead>
<tr>
<th>Rating:</th>
<th>Very Poor</th>
<th>Poor</th>
<th>Fair</th>
<th>Average</th>
<th>Good</th>
<th>Very Good</th>
<th>Superior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box</td>
<td>3</td>
<td>5</td>
<td>11</td>
<td>11</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>String</td>
<td>2</td>
<td>8</td>
<td>5</td>
<td>15</td>
<td>6</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Money</td>
<td>1</td>
<td>5</td>
<td>17</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 3
Chi-square with Context as the Independent Variable
and Teacher Ratings of Story Creativity as the Dependent Variable

<table>
<thead>
<tr>
<th>Rating:</th>
<th>Very Poor</th>
<th>Poor</th>
<th>Fair</th>
<th>Average</th>
<th>Good</th>
<th>Very Good</th>
<th>Superior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usual</td>
<td>53</td>
<td>46</td>
<td>40</td>
<td>58</td>
<td>45</td>
<td>39</td>
<td>19</td>
</tr>
<tr>
<td>Unusual</td>
<td>48</td>
<td>28</td>
<td>51</td>
<td>52</td>
<td>46</td>
<td>38</td>
<td>37</td>
</tr>
</tbody>
</table>
The fourth and final chi-square analysis also dealt with context of the story-starters. The independent variable was context of the story-starter; the categorical dependent variable was judges' scores of the stories. The judges' scores were placed in the same seven categories (very poor vs. poor vs. fair vs. average vs. good vs. very good vs. superior). It was found that the judges did not place stories in any of the seven categories more often than would be expected by chance, $\chi^2(6)=3.12, p>.05$. The data are summarized in Table 4.

**General Linear Regression Analyses**

The following results were found to be significant when analyzing the data using general linear regression analysis. When using teacher ratings of story creativity as the criterion variable, one predictor variable, total score, was found to be the most significant, $F(1,110)=67.91; p<.001$. Originality was significant, $F(1,110)=8.28; p<.01$. Elaboration, $F(1,110)=4.31; p<.05$, and irrelevant flexibility, $F(1,110)=4.34; p<.05$, were also found to be significant. See Table 5 for a summary of the regression analysis using teacher ratings of story creativity as the criterion variable. The results indicate that four of the original eight predictor variables are valid to use as an expressive language scoring system based on the teacher ratings of story creativity. The correlation between total score and teacher ratings of story creativity .67, and total score accounted for 45 percent of the variance in teacher ratings of creativity.

When using teacher ratings of student creativity as the criterion variable, only one predictor variable, total score, was found
Table 4

Chi-square with Context as the Independent Variable
and Judges' Scores as the Dependent Variable

<table>
<thead>
<tr>
<th>Rating:</th>
<th>Very Poor</th>
<th>Poor</th>
<th>Fair</th>
<th>Average</th>
<th>Good</th>
<th>Very Good</th>
<th>Superior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usual</td>
<td>3</td>
<td>8</td>
<td>18</td>
<td>18</td>
<td>10</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Unusual</td>
<td>3</td>
<td>10</td>
<td>15</td>
<td>16</td>
<td>8</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>
Table 5
General Linear Regression with Teacher Ratings of Story Creativity as the Criterion Variable

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>118</td>
<td>6305.933</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression</td>
<td>8</td>
<td>2807.004</td>
<td>350.857</td>
<td>11.03</td>
<td>.001</td>
</tr>
<tr>
<td>Total Score</td>
<td>1</td>
<td>2159.968</td>
<td>2159.968</td>
<td>67.91</td>
<td>.001</td>
</tr>
<tr>
<td>Ideational Fluency</td>
<td>1</td>
<td>15.026</td>
<td>15.026</td>
<td>0.47</td>
<td>n.s.</td>
</tr>
<tr>
<td>Associational Fluency</td>
<td>1</td>
<td>1.526</td>
<td>1.526</td>
<td>0.05</td>
<td>n.s.</td>
</tr>
<tr>
<td>Elaboration</td>
<td>1</td>
<td>137.076</td>
<td>137.076</td>
<td>4.31</td>
<td>.05</td>
</tr>
<tr>
<td>Relevant Flexibility</td>
<td>1</td>
<td>34.318</td>
<td>34.318</td>
<td>1.08</td>
<td>n.s.</td>
</tr>
<tr>
<td>Irrelevant Flexibility</td>
<td>1</td>
<td>138.019</td>
<td>138.019</td>
<td>4.34</td>
<td>.05</td>
</tr>
<tr>
<td>Originality</td>
<td>1</td>
<td>263.441</td>
<td>263.441</td>
<td>8.28</td>
<td>.01</td>
</tr>
<tr>
<td>Organization</td>
<td>1</td>
<td>57.630</td>
<td>57.630</td>
<td>1.81</td>
<td>n.s.</td>
</tr>
<tr>
<td>Residual</td>
<td>110</td>
<td>3498.929</td>
<td>31.808</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
to be significant, $F(1,110)=9.61; p<.01$. A summary of the results using teacher ratings of student creativity as the criterion variable may be found in Table 6. It appears that the scoring system is not as valuable when used to predict the classroom teacher ratings of student creativity as when predicting teacher ratings of story creativity.

**Stepwise Procedures**

A stepwise procedure was used in order to see which of the many possible combinations of the predictor variables was the best to use when predicting teacher ratings of story creativity and teacher ratings of student creativity. The best prediction model for teacher ratings of story creativity is the total score, originality and organization, which together accounted for 42 percent of the variance. No other predictors significantly contributed to the overall variance, given those three. Table 7 shows a summary of the results of the stepwise procedure using teacher ratings of story creativity as the criterion variable.

The best prediction model for teacher ratings of student creativity is total score and originality, which together account for nine percent of the variance. Given those two, no other predictors significantly contributed to the total variance in teacher ratings of student creativity. Thus, it appears that the scoring system is not a very powerful predictor of teacher ratings of student creativity. Results of the stepwise procedure using teacher ratings of story creativity as the criterion variable are summarized in Table 8.
### Table 6

General Linear Regression with Teacher Ratings of Student Creativity as the Criterion Variable

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>118</td>
<td>191.109</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression</td>
<td>8</td>
<td>24.002</td>
<td>3.000</td>
<td>1.97</td>
<td>n.s.</td>
</tr>
<tr>
<td>Total Score</td>
<td>1</td>
<td>14.595</td>
<td>14.595</td>
<td>9.61</td>
<td>.01</td>
</tr>
<tr>
<td>Ideational Fluency</td>
<td>1</td>
<td>0.565</td>
<td>0.565</td>
<td>0.37</td>
<td>n.s.</td>
</tr>
<tr>
<td>Associational Fluency</td>
<td>1</td>
<td>1.779</td>
<td>1.779</td>
<td>1.17</td>
<td>n.s.</td>
</tr>
<tr>
<td>Elaboration</td>
<td>1</td>
<td>0.616</td>
<td>0.616</td>
<td>0.41</td>
<td>n.s.</td>
</tr>
<tr>
<td>Relevant Flexibility</td>
<td>1</td>
<td>0.282</td>
<td>0.282</td>
<td>0.19</td>
<td>n.s.</td>
</tr>
<tr>
<td>Irrelevant Flexibility</td>
<td>1</td>
<td>0.174</td>
<td>0.174</td>
<td>0.11</td>
<td>n.s.</td>
</tr>
<tr>
<td>Originality</td>
<td>1</td>
<td>4.725</td>
<td>4.725</td>
<td>3.11</td>
<td>n.s.</td>
</tr>
<tr>
<td>Organization</td>
<td>1</td>
<td>1.265</td>
<td>1.265</td>
<td>0.83</td>
<td>n.s.</td>
</tr>
<tr>
<td>Residual</td>
<td>110</td>
<td>167.107</td>
<td>1.519</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7

Stepwise Procedure with Teacher Ratings of Student Creativity
as the Criterion Variable

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>118</td>
<td>191.109</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression</td>
<td>1</td>
<td>14.595</td>
<td>14.595</td>
<td>9.67</td>
<td>.01</td>
</tr>
<tr>
<td>Total Score</td>
<td>1</td>
<td>14.595</td>
<td>14.595</td>
<td>9.67</td>
<td>.01</td>
</tr>
<tr>
<td>Residual</td>
<td>115</td>
<td>176.514</td>
<td>1.509</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 8
Stepwise Procedure with Teacher Ratings of Student Creativity as the Criterion Variable

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>118</td>
<td>6305.933</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression</td>
<td>3</td>
<td>2669.530</td>
<td>889.843</td>
<td>28.14</td>
<td>.001</td>
</tr>
<tr>
<td>Total Score</td>
<td>1</td>
<td>2027.657</td>
<td>2027.657</td>
<td>64.12</td>
<td>.001</td>
</tr>
<tr>
<td>Originality</td>
<td>1</td>
<td>390.298</td>
<td>390.298</td>
<td>12.34</td>
<td>.001</td>
</tr>
<tr>
<td>Organization</td>
<td>1</td>
<td>174.920</td>
<td>174.920</td>
<td>5.53</td>
<td>.05</td>
</tr>
<tr>
<td>Residual</td>
<td>115</td>
<td>3636.403</td>
<td>31.621</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER V

DISCUSSION

Inter-judge Reliability

According to the results of the average correlation between judges, the scoring system appears to be reliable among scorers. It would be helpful to have a training session and then reestablish the inter-judge reliability of this instrument each time it is used. This would give the experimenters an opportunity to study the stability of inter-judge reliability over time.

Intra-judge Reliability

Intra-judge reliability indicated that the judges rescored the stories in much the same way as they scored them the first time. The consistency of scoring within a person across time adds power to the reliability of the scoring system.

Story-starter Content and Context

The results from performing a chi-square analysis on the content and context of the stories helped to determine if the teachers and judges categorized a greater number of stories than would be expected by chance into each of the seven creativity rating categories based on content and context. It appears that teachers were influenced by the content of the stories. This is important to know for
future research using story-starters so that content can be controlled. Judges did not place stories in any of the seven categories more often than would be expected by chance, and they were not influenced by the content of the stories. The context of the story-starters did not significantly influence teacher ratings or judges' scores, and therefore, does not need to be controlled for in the future.

These results suggest that the teachers and judges were not looking at the stories in the same way when rating and scoring them. Teachers' ratings were influenced by content whereas judges' scores were not. Though the scoring system appears to be accurately used by the judges, the scoring system does not accurately reflect teacher ratings according to content.

**General Linear Regression Analysis**

As previously mentioned, it was felt that it is essential to create a scoring system which best reflected what the teachers listed as being factors in creativity since teachers are the actual, real world judges of a student's creative and expressive language abilities. Thus, the single, independent variables which most adequately predicted the teachers' listed criteria, as determined by regression analysis, were total score, originality, irrelevant flexibility and elaboration. When each of these four factors were individually regressed on teacher ratings of story creativity, they were significantly related to the teacher ratings. Of course, total score is a combination of the seven sub-factors and includes some factors which, when used alone, are not significantly related to teacher ratings of story creativity. However, of the four sig-
significant factors, the overall total score is the most reliable and powerful. The correlation between total score and teacher ratings of story creativity was .67, and total score was the best overall predictor of teacher ratings of story creativity as revealed by regression.

Each of the students' present classroom teachers rated the students according to their "general level of creativity." General linear regression analysis showed only one significant predictor for teacher ratings of student creativity, total score.

**Stepwise Procedures**

In looking for the best combination of predictors for teacher ratings of story creativity, a stepwise procedure yielded a slightly different combination of factors than the general linear regression analyses. Of the single factors and many possible combinations of factors, it was found that the combination of total score, originality and organization accounted for the largest possible amount of variance when predicting teacher ratings of story creativity. Again, total score includes factors which are not significant in themselves but when used in combination with one another become more significant. Originality and organization also appear to contribute significantly to the overall variance. However, it is felt that in order to develop the most reliable and valid expressive language scoring system, based on teacher ratings of story creativity, total score is the best predictor. Therefore, each of the seven sub-factors of the total score need to be included in the expressive language scoring system.
A stepwise procedure showed that originality and total score, together, accounted for the largest amount of variance in teacher ratings of student creativity (nine percent). The relatively low predictions and correlations between the judges' scores and teacher ratings of student creativity could be due to the idea that classroom teachers are often poor judges of creativity (Torrance, 1972). Too often creativity gets overlooked or downplayed by class production, grades and achievement. Another possible reason for the results is that the classroom teachers, when rating the students, could have been influenced by other factors which are not included in the scoring system, such as art abilities.

**Implications**

It is felt that the expressive language scoring system proposed in this thesis is a reliable scoring system as demonstrated by the inter-judge and intra-judge reliability checks. The results showed which of the original eight variables to include in the expressive language scoring system in order to establish the most accurate scoring system and which variables were significant when predicting teacher ratings of story and student creativity. The present study supports the findings of Guilford (1968) and Torrance (1974). Creativity does appear to include factors such as originality, flexibility, elaboration and fluency. In the present study, some of these factors appeared to be stronger predictors than the others, but each factor contributed to the overall total score.

However, the scoring system is not flawless, and the author suggests that further research and development of the expressive language scoring system take place. Perhaps improvement of the scoring
system, itself, via a more in-depth analysis of the sub-factors, which may become more powerful to the total score upon revision, would be helpful. The influences of content and context on scoring would also be of interest.

The scoring system, whether revised or left as it is, offers itself as an endless research tool. As suggested in the beginning of this paper, the relationship between a student's expressive language ability, his/her school achievement and his/her IQ can be studied. It would also be possible to determine which is a better predictor of school achievement, the traditional IQ score or expressive language abilities? One could study the differences in the expressive language scores of gifted students and students with low IQ scores. Other possible areas of research are the effects of demographic variables, sex and age of subjects, and socioeconomic status of subjects on the expressive language scoring system. Two important areas to investigate are the stability of a subject's response over time and the degree to which this scoring system measures expressive language under certain testing situations. Finally, it would be interesting to analyze the differences between oral and written language using the expressive language scoring system.

In conclusion, the expressive language scoring system offers many possibilities for future studies. The present system was found to be an adequate predictor of teacher judgement of creativity. The present study is only the foundation--the development--of an instrument which will hopefully continue to stimulate further research in a very difficult and subjective area of psychology, expressive language.
REFERENCES


Guilford, J.P. Varieties of creative giftedness, their measurement and development. The Gifted Child Quarterly, 1975, 19(2), 107-121.

Guilford, J.P. and Hoepfner, R. Creative potential as related to measures of IQ and verbal comprehension. *Indian Journal of Psychology*, 1966, 41, 7-16.


## STORY-STARTER MATRIX

<table>
<thead>
<tr>
<th>Context</th>
<th>USUAL</th>
<th>UNUSUAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOX</td>
<td>When I went into the kitchen, I saw a box on the table, and ...</td>
<td>I came home from school one day and saw a box floating in front of my house, and ...</td>
</tr>
<tr>
<td>STRING</td>
<td>One day I got some string, and ...</td>
<td>I saw string growing out of the ground, and ...</td>
</tr>
<tr>
<td>MONEY</td>
<td>One day I found some money, and ...</td>
<td>One night while I was lying in bed, I heard a noise outside. When I looked out my window, I saw money falling, and ...</td>
</tr>
</tbody>
</table>
APPENDIX B
CRITERIA FOR TEACHER RATINGS OF STORIES

Teacher Number(On outside of envelope)_______

1. Read each of the stories and rate them using the following scale:
   1=worst RATING: Based on Creativity level
   2=average of the story--not mechanics
   3=best of writing!

   After each story has been placed in either stack 1, 2 or 3 . . .

2. Take stack "1" and rate each story as being either "1a" or "1b"
   using the following scale:
   1a=very poor
   1b=poor

   *Place 1a and 1b in separate stacks.

3. Take stack "2" and rate each story as being either "2a", "2b" or
   "2c" using the following scale:
   2a=fair
   2b=average
   2c=good

   *Place 2a, 2b and 2c in separate stacks.

4. Take stack "3" and rate each story as being either "3a" or "3b"
   using the following scale:
   3a=very good
   3b=superior

   *Place 3a and 3b in separate stacks.

Thus, each story has been assigned one of the following ratings
and has been placed in stacks accordingly:

1a=very poor
1b=poor
2a=fair
2b=average
2c=good
3a=very good
3b=superior

*Note: In doing this you may rearrange the stories from one
stack to another as often as you want, i.e., you may
change your initial ratings.
5. For each stack list a description of the criteria used for placing stories in that stack (paper provided).

6. Attach each list to the corresponding stack of stories. Put your seven stacks and their attached lists of criteria in the envelope provided and return to your instructor.

In order to receive your $10.00, please sign your name and social security number at the bottom of the page.

Thank You
TEACHER-GENERATED CRITERIA FOR CREATIVITY

Criteria listed for placing stories in stack 3b--Superior:

- imaginative conclusion
- creative turn, unexpected
- fully developed story
- creative use of ordinary object
- story comes alive
- develops new word, idea or thing
- story content highly unusual, unique, original, creative
- tremendous imagination required
- enjoyable to read, amusing
- different colored people in different forms
- intelligence and thought required
- fluent, "action-packed"
- abstract thinking
- audacious, vivid, colorful details
- extraordinary endings
- originality of plot

Criteria listed for placing stories in stack 3a--Very Good:

- works main idea through entire story
- may involve more than one creative idea
- extended story
- descriptive words
- new idea
- story content unusual
- considerable imagination
- well-written
- interesting but no element of surprise
- objects are flying, disappearing, talking
- original, but not as creative as 3b
- weakness found in the conclusion
- unexpected events
- using inappropriate action as a source of humor
- letting reader determine cause of action
- logical explanation for unnatural event
- giving a child or other unlikely character a position of authority
- more intricate sequence of events
- consistent unfolding of plots
- much originality and adventure
- action
- better story-tellers
- element of surprise
- holds the reader's interest
- develops some new vocabulary
- leaves room for story to be further developed
Criteria listed for placing stories in stack 2c--Good:

- conclusion expresses different idea but not creative
- different use or idea
- dialogue, optional
- extension of thought
- content not as original as with very good stories
- unrelated to main idea
- definite imagination required
- story content unusual
- some creativity but not really unique
- interesting story
- attempt to create the unique
- lack of fluency prohibits development of ideas
- some originality but too much predictability
- humor
- unexpected events creating element of surprise
- naming things and characters
- artistic outlook on ordinary objects
- use of exaggeration as found in tall-tales
- unexpected endings
- subject matter less consistent with childrens' fiction
- more specifics
- excitement, adventure
- better endings
- detailed explanations, more adjectives
- consistent
- developed story

Criteria listed for placing stories in stack 2b--Average:

- reflects material wanted
- indoctrination
- television and movies
- some imagination
- an unusual twist or touch of fantasy
- typical conclusion
- highly influenced by everyday happenings and already-learned
- story-book tales
- different or uncommon story
- some detail
- dialogue is replacing creative action
- life situations, commonplace objects
- use of personification
- unexpected action
- associating meaning with color
- time lapse used intentionally
- better plotting, more body to the story
- catchy ending
- easy to follow
- resolved the original idea, conclusion
- logical explanations
Criteria listed for placing stories in stack 2a--Fair:

- fair story, not very creative, some imagination
- slight tie-in to starting line
- story has an unexpected twist or touch of fantasy
- no new ideas
- readable
- common theme but makes effort towards unique twist
- story not based on common information
- attempts something unusual
- too true-to-life to be entertaining
- more happening than with 1b, better ideas
- attributes unexplainable to dreams
- stories end "happily ever after"
- more conversation used
- more action than in 1b
- better thought out than 1b
- some surprise endings
- doesn't describe in much detail
- not very consistent
- jumbled thinking

Criteria listed for placing stories in stack 1b--Poor:

- hard to read
- difficult to understand
- little imagination
- indication that the event might actually happen
- hard to follow, loose ideas
- mainly wrote about an everyday occurrence
- t.v. influence
- personal experience
- explained the "why" of the story
- completion of the story
- some imagination concerning completion or detail
- too much attention to detail and dialogue
- too practical in content
- probable use of ordinary things
- inability to make a decision as required by the topic
- poor or inconsistent endings
- a lot of filling in
- trite
- not enough detail

Criteria listed for placing stories in stack 1a--Very Poor:

- makes little or no sense, got off the subject
- no point (beginning doesn't relate to the end)
- no imagination
- actual event in past, everyday occurrence
- hard to read
- average responses to what Mom and Dad told them to do
Criteria listed for stack 1a--continued:

doesn't write or follow instructions
known material used
no creativity in developing a theme
didn't explain the "why" behind the story
drew ideas from fairy tales
too practical in content
concluding story before a logical conclusion was reached
platonic descriptions
repetitions
monotonous reading
very little use of descriptors
DIRECTIONS FOR USING THE EXPRESSIVE LANGUAGE SCORING SYSTEM

I. Ideational Fluency

Count the number of unique, different uses of the main object--either box, string or money--within the story. Think of the use as "serving a purpose". The use can be implied, as with example 2.

II. Associational Fluency

Count the number of things done "with", "to" or "by" each object and/or the consequence of the usage. What have you done with the object, what are the uses to which it is put, what was done to it if it was transformed, i.e., changed?

**You can have one--ideational fluency or associational fluency--without the other.

**If counted once, don't count again, e.g., I took the money home and took it to my room--Associational fluency=1, took

**Conditionals--could, should, wanted to--don't count, e.g., I could catch it--Associational fluency=0

**Past tense or present tense does count, e.g., I made it--Associational fluency=1, made

**If the associational fluency pertains to a necessary or important part of the object, but not necessarily the whole object, it is still counted, e.g., I took the top off of the box--Associational fluency=1, took

Examples:

I. A.

1. I took the box and put my toys in it then I decided to use the box as a hamster cage. 2 0

2. The box had candy in it. I (put) the box in my room. 1 1
   --this implies that the box was used as a container.

3. The money was from Mexico. I (hid) it in my room. 0 1

4. One day I got some string and made a kite tail for my kite. The tail (broke) and my kite fell. 1 1

5. I (hid) the money in my drawer. Later I bought an ice-cream cone with it. 1 1
I. A.

6. "...When I looked out my window, I saw money falling..." and it fell into our swimming pool. I (dove) for pennies the next day and (put) them in a jar. Then I (counted) them. 0 3

7. "I saw string growing out of the ground and..." I cut the string into pieces and used them to (tie) up my hair. 1 1

8. ...I cut the string into pieces and used them to (tie) up my hair and to (lace) up my shoes. 1 2

III. Elaboration

Count the number of descriptors and qualifiers which give power or aid in clarification/understanding of the story. If the description is needed to complete a thought, it is not counted. Details over and above those necessary to communicate the basic idea are counted.

**Prepositional phrases are not counted if they are the object of the verb, e.g., The man is in the store--Elaboration=0

**Prepositional phrases are counted if they are not necessary to complete the thought, e.g., The box floated out the door--Elaboration=1, out the door

**Verbs are not counted as descriptors.

**Conjunctions only count if they start off a sentence, e.g., But he ate it--Elaboration=1, But

**Phrases within a sentence that start with because or so count as elaboration, e.g., He ate it because he was hungry--Elaboration=1, because he was hungry

**Don't count possessive pronouns or articles, e.g., My house is the house that burned--Elaboration=0

**Qualifiers such as maybe, instead, then, finally, and at last count.

**Redundancies count only once--see example 11.

Examples:

9. The blue box was full of bubble gum. 3

10. While I was eating, a robber came in and took the money. 1
11. Finally, I bought the candy and ate and ate and ate.  
   Elab. 2

12. The money was on the table.  

13. A mean man came and grabbed the box out of my hand. He ran too fast/ for me/ to catch him.  
   Elab. 6

14. The box was heavy.  

IV. Relevant Flexibility

Count the number of relevant changes or shifts in the focus or approach of the story. Think of the shifts as being a general change in the direction of the story. Divide by ideas or sub-themes within the story.

V. Irrelevant Flexibility

Count the number of irrelevant changes or shifts in the focus, or approach, of the story and subtract from the total.

**If a sentence or sentences refer back to a previous idea, it is not counted again.

Examples:

15. I took the money to the store and bought a new dress. I hope it's going to be sunny tomorrow.  
   R. I. 0 1

16. I took the money to the store and bought a new dress. Tomorrow I'm going to wear the dress to school.  
   R. I. 1 0

17. "I came home from school one day and saw a box floating in the air in front of my house and..." the box floated all over town. I followed it everywhere it went until it stopped at the baseball park. There, I saw a bunch of my friends playing ball, and I played with them until dark. By the time I walked home, it was time to eat supper.  
   R. I. 3 0

18. "I saw string growing out of the ground and..." it grew to be as big as a flag pole. I wanted to try and climb it. Last year I climbed a mountain.  
   R. I. 1 1

19. The box was taped up so good that I couldn't even peek in it. I needed x-ray vision. Last year I got glasses. They are black. When mom got home, she let me open the box.  
   R. I. 1 1
20. I tied the string into a thousand knots. My hands were tired from all the knots I tied. The next day I had blisters on my hands. (tieding knots implies the consequence of blisters)

21. We had fun trying to catch the box. It flew faster than I could run. My favorite t.v. show was on last night.

VI. Originality

Originality is defined as the uniqueness of the main, overall usage of the object. The full story was read in order to determine the primary usage or theme. The stories were divided into their six categories and the primary usage was determined for each story within its group. Responses--usages--which were found in only one story out of the possible 20 received a score of 4. Responses found within two stories received a score of 3. Responses found within three stories received a score of 2. Responses found within four stories received a score of 1, and responses found within five or more of the stories received a score of 0.

**If, however, the response was a transformation, i.e., the object was taken and changed from a "typical" usage--regardless of where it fell in the above categories, it received an additional bonus point. Thus, a response which was found within five or more stories but involved a transformation received a score of 1.

**An additional bonus point was given for a creative twist, i.e., a surprise ending.

Thus, a subject's originality score can vary between 0 and 6.

Examples of transformations:

- box changed to a fort
- string changed into a bracelet
- money changed into a book-marker

Examples of "typical", expected usages of the objects:

- box--holding something, a container
- string--tie things up
- money--buy something, to spend

VII. Organization

Beginning with the story-starter, count the number of sentences which are related to the prior sentence. If a sentence is related to the sentence prior to it, it receives a point. If a sentence is not related to the prior sentence, it does not
receive a point. Judge the relationship by asking, "Is the idea of the sentence related to the prior sentence?"

The score equals the total number of points received for related sentences divided by the total number of sentences (**see below) multiplied by 10. The story-starter does not count as a sentence, but the first phrase or sentence written by the subject must be evaluated for relationship to the story-starter.

****Before scoring, break the run-on sentences and the compound sentences into separate sentences in order to obtain the total number of sentences in the story.

**When judging for relationship to the prior sentence, take off or omit the beginning qualifiers, e.g., then, finally, instead, so **Completion of the story-starter by the subject counts as one sentence.

Examples:

R.S. = related sentence  
S. = total number of sentences  
Sc. = final score

<table>
<thead>
<tr>
<th>R.S.</th>
<th>S.</th>
<th>Sc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>22.  &quot;One day I found some string and...&quot; I went home and ate supper.</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>23.  &quot;One day I found ...&quot; and I made a kite.</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>24.  &quot;As I came home from school I saw a box floating and...&quot; I ran in and changed my clothes. I put on my yellow shirt. Then, I went out and tried to catch the box. It was too quick for me.</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>
APPENDIX E
# INTRA-JUDGE RELIABILITY TABLE

<table>
<thead>
<tr>
<th>Judge:</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideational Fluency</td>
<td>.70</td>
<td>.85</td>
<td>.95</td>
<td>.90</td>
<td>.80</td>
</tr>
<tr>
<td>Associational Fluency</td>
<td>.85</td>
<td>.70</td>
<td>.70</td>
<td>.75</td>
<td>.69</td>
</tr>
<tr>
<td>Elaboration</td>
<td>.85</td>
<td>.81</td>
<td>.87</td>
<td>.92</td>
<td>.87</td>
</tr>
<tr>
<td>Relevant Flexibility</td>
<td>.83</td>
<td>.88</td>
<td>.86</td>
<td>.89</td>
<td>.77</td>
</tr>
<tr>
<td>Irrelevant Flexibility</td>
<td>.90</td>
<td>.80</td>
<td>.80</td>
<td>1.00</td>
<td>.60</td>
</tr>
<tr>
<td>Originality</td>
<td>.98</td>
<td>.95</td>
<td>.90</td>
<td>.96</td>
<td>.96</td>
</tr>
<tr>
<td>Organization</td>
<td>.77</td>
<td>.86</td>
<td>.85</td>
<td>.81</td>
<td>.93</td>
</tr>
<tr>
<td>Total Score</td>
<td>.89</td>
<td>.90</td>
<td>.93</td>
<td>.93</td>
<td>.92</td>
</tr>
</tbody>
</table>