2001

A study of metaphor development in young gifted children

Catherine Anne Little
William & Mary - School of Education

Follow this and additional works at: https://scholarworks.wm.edu/etd
Part of the Curriculum and Instruction Commons, Developmental Psychology Commons,
Elementary Education Commons, Language and Literacy Education Commons, and the Special
Education and Teaching Commons

Recommended Citation
Paper 1550154117.
https://dx.doi.org/doi:10.25774/w4-smgv-h760

This Dissertation is brought to you for free and open access by the Theses, Dissertations, & Master Projects at W&M ScholarWorks. It has been accepted for inclusion in Dissertations, Theses, and Masters Projects by an authorized administrator of W&M ScholarWorks. For more information, please contact scholarworks@wm.edu.
A STUDY OF METAPHOR DEVELOPMENT
IN YOUNG GIFTED CHILDREN

A Dissertation
Presented to
The Faculty of the School of Education
The College of William and Mary in Virginia

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy

by
Catherine A. Little
September 2001
A STUDY OF METAPHOR DEVELOPMENT
IN YOUNG GIFTED CHILDREN

by

Catherine A. Little

Approved September 2001 by

Joyce VanTassel-Baska, Ed.D
Chair of Doctoral Committee

James H. Stronge, Ph.D.

Robert J. Hanny, Ph.D.
# TABLE OF CONTENTS

Acknowledgments..........................................................................................................v

List of Tables ..................................................................................................................ix

List of Figures ................................................................................................................xi

Abstract ..........................................................................................................................xii

Chapter I: The Problem..................................................................................................2
  Introduction........................................................................................................2
  Statement of the Problem ...................................................................................10
  Theoretical Assumptions ...................................................................................12
  Purpose and Research Questions .......................................................................14
  Rationale ............................................................................................................15
  Summary of Methodology .................................................................................16
  Delimitations and Limitations ...........................................................................17
  Definitions of Related Terms.............................................................................19

Chapter II: Review of the Related Literature.................................................................26
  Introduction........................................................................................................26
  Strand I: Early Literacy Development and Instructional Trends in Reading.....28
  Strand II: The Nature of Metaphor and the Development of Figurative
             Competence .............................................................................................34
  Strand III: Young Gifted Children and Figurative Competence....................55
  Strand IV: Figurative Competence and Instruction ...........................................62

Chapter III: Methodology ..............................................................................................74
  Purpose and Research Questions .......................................................................74
  Conceptual Framework ......................................................................................75
  Research Design .................................................................................................77
  Sample ................................................................................................................80
  Instrumentation ..................................................................................................84
  Study Intervention Procedures ..........................................................................99
  Data Collection ................................................................................................105
  Data Analysis ....................................................................................................106
  Delimitations and Limitations ..........................................................................107
  Ethical Considerations .......................................................................................109
Chapter IV: Findings........................................................................................................111
  Introduction...........................................................................................................111
  Findings Related to Research Question 1............................................................111
  Findings Related to Research Question 2............................................................119
  Findings Related to Intervention Effects.............................................................145
  Summary of Findings...........................................................................................175

Chapter V: Discussion, Conclusions, and Implications...................................................179
  The Development of Figurative Competence......................................................179
  The Effects of Instruction ....................................................................................186
  Conclusion ...........................................................................................................193
  Implications for Research and Practice...............................................................194
  Summary .............................................................................................................201

References.......................................................................................................................203

Appendix A: Demographic Data on Schools.............................................................217
Appendix B: Metaphor Comprehension Test .............................................................218
Appendix C: Literary Analysis and Interpretation Test...............................................224
Appendix D: LAIT Scoring Rubric ...........................................................................226
Appendix E: Teacher Assessment of Student Performance.......................................231
Appendix F: Classroom Observation Form ...............................................................233
Appendix G: Curriculum Framework and Lesson Outline.........................................234
Appendix H: Qualitative Analysis of Metaphor Production......................................237
Appendix I: Qualitative Analysis of Selection of Vehicles.......................................241
Appendix J: Demographic and Implementation Questions Asked of Treatment Teachers .........................................................................................................................243

Vita .............................................................................................................................244
ACKNOWLEDGMENTS

It seems only fitting, in the preliminary pages of this study of metaphor, to use a figurative expression to exemplify what this process has meant. Jane Yolen, whose works are used extensively in this study’s intervention, wrote the following: “Each day is a journey, a leaving home.... Only if you look will you find. Only if you leave will you arrive. One step, then another, as day unrolls itself along the road toward night.” The process of conducting this study has represented a long journey for me, perhaps most challenging in its requirements that I find the courage to look, to leave, to take the steps beyond home into uncertainty. I could not have made the journey without the many companions who offered shelter, pointed the way, or just walked along beside me.

First I must acknowledge the faculty and students of the School of Education at the College of William and Mary. Their continued guidance and fellowship have offered challenge and stimulation to my mind and heart, encouraging the type of reflective practice that characterizes the best in educational professionalism. The investment of the faculty in the overall learning experience of students has been an inspiration, and their wealth of knowledge and experience has represented an invaluable resource. I thank each of them for the learning experiences they have provided, and I am grateful to Dean Virginia McLaughlin for her leadership in fostering the environment of collaborative learning that welcomes participation at every stage of educational experience.

Among the faculty, there are several I wish to acknowledge individually for their assistance with this program and this study. Dr. Robert Hanny has been a quiet, guiding force throughout my entire experience at William and Mary, from the time I walked into
his office as a freshman advisee, through courses and conversations across my graduate program, to the momentous days of proposal and dissertation defense. Dr. Hanny’s deep understanding of the endeavor of curriculum, instruction, and assessment has had an important influence on my own developing emphasis in this area of study, and his willingness to sit down and discuss ideas has stimulated me on the good days and inspired me on the bad days to keep working and keep thinking. Dr. James Stronge’s unfailing dedication to the teaching and learning enterprise and his commitment to discovering what makes teaching work have helped to keep me grounded in what matters in this field. An inspirational teacher himself, he models critical thinking and the need to challenge students in every question he asks. His generosity in terms of time, flexibility, and moral support have been invaluable throughout the process of developing this study. Both of these gentlemen enriched my study by agreeing to serve on my committee.

The most powerful educational influence for me in this program and in this study has come from my chair and advisor, Dr. Joyce VanTassel-Baska. I came to William and Mary for my Master’s degree because I knew William and Mary; I stayed for a doctorate because of Dr. VanTassel-Baska. She exemplifies scholarship, with a prolific publication record and ever a focus on new questions to be asked from every answer she finds, and she is also one of the most effective teachers I have ever known. Part of her teaching talent is grounded in her ability to understand the needs of every learner; as an advisor, this ability also allows her to know for each student when to challenge and when to accept, how far to push and how often to praise. In developing my dissertation, there were many times when I was prepared to quit, and were many more times when I procrastinated and postponed, but Joyce’s faith in me persisted and helped me find the vi 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
strength to continue. I have been honored by my many opportunities to work with her and learn from her; she has made me the teacher and learner that I am.

Many other individuals also deserve acknowledgment for their support in this process. Virginia Tucker, Colleen O’Donnell, and Sherry Watts of Montgomery County’s Division of Accelerated and Enriched Instruction provided me with access to teachers and students and helped me tremendously with details of implementation. The teachers in the study group were, of course, vital to the process; I thank each and every one of them for their willingness to participate and for their hard work and enthusiasm throughout the project. I also thank Williamsburg-James City County Schools for its involvement as a pilot and comparison site. And of course, I thank all of the second graders who participated for their involvement and for their beautiful metaphors.

My colleagues at the Center for Gifted Education past and present have provided moral support and practical help throughout the process. I particularly wish to acknowledge a few of them: Dr. Linda Avery, who was the first to read every chapter and boosted my ego before I had to turn pages in; Dr. Li Zuo, whose guidance on statistical matters was invaluable and moral support even more so; Jeanne Struck, who spent hours scoring tests with me; Dr. Jane Bailey, who reviewed the unit and offered considerable guidance regarding metaphor development; and Dana Johnson, who reviewed the unit at many stages and is ever a pillar of strength and support. I also thank the many friends, colleagues, and fellow students whose willingness to listen to my laments on some days and elation on other days was greatly appreciated, because it made the way less lonely, especially Dr. Elissa Brown, Diann Drummond, and Molly Sandling.

vii
As this part of my life’s journey reaches its conclusion, I must above all recognize those who have traveled with me longest. My family shed blood, sweat, and tears along with me over this process, and I thank God every day for each one of them. Even though on the hardest days I begged my father not to ask how the dissertation was going, his unfailing interest in my progress was always comforting as much as it was challenging. My mother has probably learned more about educational research than she ever wanted to know from the hours I have spent telling her about the study, but she is always willing to listen. Throughout my education, my parents have alternately commiserated with me about how much work I had to do, cajoled me into doing some of it, and celebrated with me when milestones are achieved, and I am forever grateful for their support. My sisters, too, have had an incalculable influence on my progress. My sister Jennifer first gave me the idea to pursue gifted education, and I thank her for never letting me give up on my dreams and for having such faith in me that I always work harder to try to live up to what she sees. My sister Jeannine shared her hospitality with me when I traveled to visit my study teachers, making the trips more enjoyable and reminding me of the need for balance in my life even through these times of stress. She also sat beside me for many hours of work on this project and prevented me from leaving too much of it until the last moment. I thank them both from the bottom of my heart.

Metaphors of traveling abound in graduate school; there are lights at the end of tunnels, paths to follow, and forks in the road. As I reach the end of this part of the journey, I know there are many more to come, for “way leads on to way.” I am deeply grateful to all those who have traveled with me this far.
LIST OF TABLES

1. Summary of Studies Relating Ability Measures to Figurative Competence ............56
2. Research Questions and Relevant Data Sources and Analyses ................................79
3. Schools Included in Sample ......................................................................................81
4. Item Specifications for Metaphor Comprehension Test ...........................................86
5. Equivalent Forms Pilot Results (MCT) .....................................................................88
6. Equivalent Forms Pilot Results (MCT), Outliers Removed .....................................89
7. Form A Test-Retest Reliability (MCT). ....................................................................90
8. Form B Test-Retest Reliability (MCT) ....................................................................91
9. Means of LAIT Form Scores ....................................................................................95
10. Means of LAIT Form Scores, Outlier Removed ......................................................96
11. Form A Test-Retest Reliability (LAIT) .................................................................96
12. Form B Test-Retest Reliability (LAIT) ....................................................................97
14. Comparison of TCS Subtest Results for Treatment and Comparison Groups...........112
15. Descriptive Results on MCT ................................................................................114
16. Regression Analysis Summary for TCS Subtests Predicting MCT Gain Scores ......115
17. Regression Analysis Summary for TCS Subtests Predicting MCT Posttest Scores...116
18. Descriptive Results on LAIT ................................................................................116
19. Regression Analysis Summary for TCS Subtests Predicting LAIT Gain Scores.....117
20. Regression Analysis Summary for TCS Subtests Predicting LAIT Posttest Scores .118
21. Comparative Statements from Treatment Classes ...................................................120
22. Quantitative Summaries of Comparative Statements .............................................121
23. Summary of Selection of Vehicle Responses ...........................................................134
24. Classifications of Metaphor Interpretations ..............................................................141
25. Number and Percentage of Required Lessons Implemented by Treatment Group Teachers........................................................................................................147
26. Classroom Observation Scores for Treatment Group Teachers .........................152
27. Analysis of Covariance for Treatment and Comparison Groups on MCT ..........155
28. Analysis of Covariance for Treatment and Comparison Groups on MCT, One Class Removed...................................................................................................156
29. Pretest-Posttest Correlations by Item Type – MCT .............................................159
30. Item Analysis – MCT ..............................................................................................160
31. Descriptive Statistics for Gender on MCT (Treatment Group) ............................162
32. Pretest and Posttest Means by Class Group – MCT ..............................................163
33. Analysis of Covariance of Treatment and Comparison Groups on LAIT ............164
34. Pretest-Posttest Differences on LAIT in Treatment and Comparison Groups .....165
35. Analysis of Covariance by Gender on LAIT (Treatment Group) ........................166
36. Pretest and Posttest Means by Class Group – LAIT .............................................166
37. Item Analysis – LAIT .............................................................................................167
38. Mean TASP Scores in Treatment Classes ..............................................................170
39. Item Correlations Between TASP and Item Groups on MCT ..............................172
40. Correlation Coefficients for Relationships of TASP Items 7 and 8 to Test Items.....174
LIST OF FIGURES

1. The Model of Figurative Competence .................................................................44
2. Metaphor Chart ...................................................................................................102
3. Literature Web ....................................................................................................103
A STUDY OF METAPHOR DEVELOPMENT IN YOUNG GIFTED CHILDREN

ABSTRACT

The purposes of this study were to explore metaphor development in young, verbally talented learners and to assess the effects of instruction on related tasks. A quasi-experimental design was used, with a sample of 70 second graders in the treatment group and 21 in the comparison group. Treatment classes engaged in a literature-based intervention focused on metaphor structure. Both groups were assessed before and after the intervention period on two metaphor instruments, with student products and teacher assessments collected from treatment classes. Data sources also included classroom observations and scores on the Test of Cognitive Skills (TCS).

Findings around metaphor development supported previous research, demonstrating that verbally talented second graders are developmentally capable of some metaphor tasks. Some products exceeded age expectations, supporting predictions that gifted students would surpass age peers. Small positive significant relationships ($p < .05$) were found between TCS scores and study instrument scores.

Findings around treatment effect were limited. On one instrument, the treatment group outperformed the comparison group but without showing growth. On the other, the treatment group showed growth not attributable to the intervention. Class-level comparisons indicated possible differences in performance related to degree of implementation of the intervention. Teacher assessments demonstrated a significant relationship ($p < .001$) to student performance on one instrument.
The study supported and extended previous research on metaphor development, with results meeting and exceeding age expectations. Intervention results were weaker, and limitations of sample size and instrumentation suggest the need for further investigation. Student products and teacher responses indicated potential appropriateness of the intervention, but further research is necessary to determine effectiveness.

CATHERINE A. LITTLE

PROGRAM IN EDUCATIONAL POLICY, PLANNING, AND LEADERSHIP

THE COLLEGE OF WILLIAM AND MARY IN VIRGINIA
A STUDY OF METAPHOR DEVELOPMENT IN YOUNG GIFTED CHILDREN
CHAPTER I
The Problem

Introduction

Learning is a process every human engages in throughout life. Broadly defined, learning involves the acquisition of new knowledge or skills. It also involves relating that new information in some way to what is already known and understood. Metaphor is one of the ways in which we relate that knowledge which we already possess to the unknown, to help our understanding of new concepts. The word metaphor itself derives from the Greek *metaphora*, meaning to transfer or to carry across, and its English meaning encompasses an intention to convey, or to transfer, a clearer or fresher meaning by use of a figure of speech (Steinbergh, 1999). Indeed, many of our fundamental conceptions of the world are based on comparisons from one object or system to another through metaphorical and analogical thought and language (Lakoff & Johnson, 1980). Moreover, metaphor is one of the enriching features of language, a device that offers opportunities to make language more resonant to the senses and inspiring to the mind’s vision.

The importance of metaphor development as an aspect of overall literacy development has been addressed by a number of psycholinguists and educational researchers. Levorato (1993) asserted that the highest level of language development is a level of metalinguistic awareness, at which point an individual is able to reflect upon language itself and manipulate it through its many purposes and meanings to achieve sophisticated communication. The prevalence of figurative language in everyday speech – to the extent that as many as one in eight utterances contains a nonliteral or figurative reference (Nippold, 1985) – suggests that the ability to comprehend and produce
metaphorical language is an important component of this overall metalinguistic awareness. Indeed, Steinbergh (1999) noted that poetry study, with emphasis on metaphor, “advances students’ control over language and increases their ability to read for both meaning and literary technique” (p. 324).

A metaphor by its nature is a tool for meaning-making; it provides comparisons across concepts from different domains, thereby allowing the reader or listener to develop a deeper understanding of the concepts addressed and to create new connections across conceptual categories; production of metaphor allows a more advanced demonstration of these same categorization and connection skills. These processes of meaning-making through language comprehension and production are the very basis of literacy development, which is itself the foundation of much of the educational process (Tierney, 1991). Because of the value of metaphor as a meaning-making tool and because of its prevalence in everyday speech and writing, the process of metaphor development in language learning and the role of metaphor instruction in language arts curriculum are important topics for consideration by educators and researchers in child development.

Educational experts and developmental psychologists have given attention to the appearance of metaphor production and comprehension in children and adults, studying the conditions under which metaphors are best understood and in which spontaneous use of metaphor appears (e.g., Gardner & Winner, 1978; Johnson, 1991, 1996; Rummel & Dykstra, 1983; Vosniadou, 1987). The research consistently demonstrates that facility with metaphor is a developmental process, manifesting itself to different degrees at different stages of the childhood years; studies have also shown that metaphor development is related to a complex of other factors, including domain-specific
knowledge, linguistic ability, reading and writing fluency, and to the context in which a
given metaphor task is presented (Broderick, 1991; Johnson, 1991; Levorato, 1993;
Vosniadou, 1987). The processes involved with analyzing metaphors include analogical
reasoning and categorization of concepts (Castillo, 1998; Vosniadou, 1987), both
advanced skills in cognitive development and useful in a myriad of learning
opportunities: “Metaphorical thinking... allows children to use existing knowledge to
understand new phenomena, phenomena that are not quite similar to anything they have
experienced before” (Vosniadou, 1987, p. 882).

Most of the research on metaphor development, however, has been laboratory-
based; children have been brought into the research setting and tested as to their level of
understanding or production of various forms of metaphors. The purpose of this type of
research around metaphor comprehension and production has been to determine
developmental patterns and to clarify the conditions under which children’s level of
development could most clearly be assessed (Levorato, 1993; Vosniadou, 1987). Because
the focus was on the developmental levels children had already attained by the time they
reached the different ages studied, less investigation has been made of the effects of
intervention on this developmental process. Several more recent studies introduced an
instructional aspect to the examination of metaphor development, exploring the effects of
instruction around specific types of figurative language or processing on young children’s
performance on metaphor tasks (Castillo, 1998; Redmond, 1997; Rudden, 1994).
However, the research as to the relationship of instruction to the overall process of
metaphor development is still limited, as is the literature on the role of metaphor
instruction in language arts curriculum in the early elementary grades.
Vygotsky (1978, 1986) theorized that individuals have two levels of development within a domain of learning: the level at which they can achieve independently, and the level at which they can achieve with carefully designed cognitive and metacognitive mediation. The distance between these two levels, in Vygotsky’s theory, is called the “zone of proximal development,” and it represents the area in which learning actually occurs. Vygotsky (1986) also noted that different children’s zones of proximal development are of different ranges, indicating that some can gain more from a given amount of instruction than others. Csikszentmihalyi (1990) has also discussed the differential zones in which individuals can achieve, calling the area between boredom and frustration “flow” and emphasizing that flow for an individual depends on both ability and experience in the given domain. Related research has also demonstrated a preference among gifted students for the types of activities that challenge them sufficiently to reach this state of flow (Csikszentmihalyi, Rathunde, & Whalen, 1992).

In the area of metaphor development, as indicated above, numerous studies have shown even preschool-aged children to be competent at responding to metaphor comprehension and production tasks, especially if presented in a context that does not require them to respond in complex linguistic utterances themselves and provided that at least the vehicle term in the metaphor is familiar (Vosniadou, 1987). Nonetheless, a key challenge of metaphor interpretation is the recognition of the disparate conceptual categories to which a metaphor refers and the implicit or explicit ground connecting those categories in a nonliteral way; thus, concept development plays a key role in conscious interpretation and production of metaphor. Vygotsky (1978) and Piaget (1952; Piaget & Inhelder, 1968) both discussed the idiosyncratic conceptual categories young children
develop based upon their experiences and their subsequent efforts to clarify these categories through linguistic trial-and-error. More recent research has borne out this tendency to develop experienced-based knowledge constructs that may, in fact, be misconceptions (Nuthall & Alton-Lee, 1993). Related recommendations with regard to effective teaching and learning practices suggest that direct and focused teaching of concepts, with related opportunities for students to apply the processes involved with concept development, can assist students in developing less idiosyncratic perspectives on the world around them and more useful tools for future learning (Taba, 1962; Vygotsky, 1986). Furthermore, according to Vygotsky (1986), the key to solidifying scientific (or theoretical) versus spontaneous (or experience-based) concepts is to clarify those concepts in language through interaction with the teacher and with other students. Metaphor, then, itself provides an excellent vehicle for using language to demonstrate both differences and similarities between conceptual categories, thus sharpening a child’s focus on the differential definitions of the concepts involved. As Steinbergh (1999) noted, metaphor can serve as a means of encouraging students to “express more abstract ideas and relationships” (p. 324).

For the verbally gifted child in the early elementary grades, the language arts curriculum can pose particular problems even as it has the potential to offer particular opportunities. Advanced vocabulary and early reading comprehension are frequently listed as indicators of potential giftedness in the verbal area (Lewis & Louis, 1991; VanTassel-Baska, Johnson, & Boyce, 1996). Oral language development is expected to progress at a rapid rate in the preschool years among most children, but some also develop early literacy or precocious reading skills as well, bringing them to the early
elementary grades with the basic reading skills of phonological decoding and sight word recognition already at an advanced level (Jackson & Klein, 1997). However, not all precocious readers demonstrate later advancement in verbal knowledge and reasoning ability; many may later be assessed within the population of students as above-average but not necessarily qualified for identification for gifted programs (Jackson, 1992; Jackson, Donaldson, & Cleland, 1988). Because of this factor of complicated predictability of future verbal abilities, along with funding constraints and real and perceived philosophical conflicts between gifted education and early childhood education (Barbour, 1992), children are frequently not identified for gifted programs until third grade or later (Kames & Johnson, 1987; 1991).

Since differentiation of instruction for gifted students in regular classroom settings is frequently limited at best (Archambault et al., 1993; Westberg, Archambault, Dobyns, & Slavin, 1993), for young gifted children the consequence of later identification is that they receive little or no differentiated service in the early elementary grades, especially since research-based materials to support differentiated instruction for this age-level population are limited (Barbour & Shaklee, 1998). This indicates a need for attention to how curriculum for young children should be planned so that it has a stronger potential for differentiation. In the area of language arts, instruction in the early elementary grades frequently places central emphasis on fluency development, either through intensive skill teaching or literature immersion, as discussed below. A primary intent of this focus on fluency is to support the allocation of cognitive capacity to comprehension rather than to the decoding process (Nathan & Stanovich, 1991). However, with such an instructional focus on fluency leading thence to comprehension, a
child who is already reading fluently at an advanced level independently and
demonstrating comprehension of grade-level texts is not likely to be working in his or her
zone of proximal development and facing challenging school activities. Thus, language
arts instruction in the primary grades must be examined for how it may challenge and
enhance the development of advanced readers.

Reading research over the past two decades has explored a variety of approaches
to the teaching of reading, including focus on phonological skills at one end of the
spectrum and a “whole-language” approach at the other end (Dahl, Scharer, Lawson,
Grogan, 1999; Daniels, Zemelman, & Bizar, 1999), with the most recent emphases on the
complexity of literacy instruction and the importance of a literacy curriculum that embeds
reading within a larger social context, with alphabetic principles and comprehension as
The “emergent literacy” school of thought also asserts that literacy develops as an
integration of reading and writing skills and involves the child as meaning-maker
throughout the process – emphasizing the development of metalinguistic awareness and
concept construction and their crucial roles in achieving literacy, as well as the
environmental factors that influence literacy development (Sulzby, 1991; Whitehurst &
Lonigan, 1998). This philosophy of reading development that involves the child as
meaning-maker and emphasizes the role of environmental stimuli in the developmental
process suggests that metaphors, which require meaning-making, may hold potential as
powerful environmental stimuli for the development of literacy. Moreover, because
emergent literacy focuses on meaning more than sight vocabulary and phonological
skills, it suggests that for precocious readers whose phonological and decoding skills are
already advanced, metaphorical language which requires careful conceptual organization may be a useful source for challenging further metalinguistic development. Thus, attention to metaphor in reading study may serve as a useful area for curricular differentiation for young, verbally advanced children.

Beyond the population of young gifted children, curriculum for gifted children in general is a major area of research and development within the field of gifted education. Studies of classroom practice have demonstrated little differentiated instruction occurring in regular elementary classrooms to serve the needs of gifted students (Archambault et al., 1993; Westberg et al, 1993), and there has been a call nationally for the development of stronger programs for gifted students in and out of the regular classroom (U.S. Department of Education, 1993). With an awareness of this situation, experts in gifted education have worked to develop curriculum specifically tailored to the needs of gifted students, with the intent that teachers who chose to use the curriculum in the regular classroom would then be meeting gifted students’ needs more effectively. The Integrated Curriculum Model (VanTassel-Baska, 1986, 1995) is a basis for curriculum development that targets the learning needs of gifted children through the three dimensions of advanced content, higher level processes and products, and abstract, interdisciplinary concepts. This model has served as the foundation for several series of curriculum units, including four units in the language arts that have been found to have significant effects on elementary and middle school gifted student performance in literary analysis and interpretation and in persuasive writing (VanTassel-Baska, Johnson, Hughes, & Boyce, 1996; VanTassel-Baska, Zuo, Avery, & Little, in press). Some of the results of the more recent study included significant gains for third grade gifted students; however,
curriculum for younger gifted children based on this model has not yet been systematically explored.

Statement of the Problem

Although researchers have devoted considerable energy to understanding metaphor development and figurative competence in children of various ages, fewer studies have been conducted to determine the potential effects of instruction on metaphor comprehension and production in children. Rather, in many cases, the metaphor studies have been laboratory-based studies that take a snapshot of children’s figurative competence at a given time and draw conclusions based on that information. The few studies of the effects of instruction that do exist have demonstrated positive effects of a minimal amount of instruction on figurative competence (Castillo, 1998; Rudden, 1994), suggesting a promising direction for further research, especially in a broader range of school settings over a longer period of time, and for targeted curriculum development.

Another area that has been given only limited attention in the research on metaphor development is the variability in performance among children at a given age based on their differences in cognitive ability, either globally or within the verbal domain. Among those studies that have examined ability difference under some definition, inconsistent results have been found. A study that separated IQ from developmental mental capacity found that the latter was significantly related to performance on tasks of metaphor interpretation, but not IQ (Johnson & Pascual-Leone, 1989). However, studies related to analogical reasoning have demonstrated performance differences between students based on their ability level as measured on standardized tests (Caropreso & White, 1994; Nippold, Martin, & Erskine, 1988). Moreover, none of
these studies involved instructional intervention. Because of the dearth of strong programs for young gifted children in schools, attention to research and development around curriculum designed to meet their needs is an important direction for the field of early childhood gifted education.

The problem of this study was to determine whether a literature-based instructional program designed to provide scaffolding for the processes of metaphor comprehension and production could affect children's performance on measures of figurative competence. The study compared differences in children's performance on several measures of figurative competence before and after participation in the instructional intervention. Additionally, because of the oral interactions and discussions that represented a central element of the intervention, teachers were asked to assess student performance across the activities of the unit as another basis for determining figurative competence. Performance on pre- and post-assessments of figurative competence among these students were compared to scores of a comparison group of children who did not participate in the program but who received the standard literature-based reading instruction for their school district across the same period of time. In addition, the study examined how children's performance on the measure of figurative competence related to their performance on standardized measures of general and verbal ability, and student products related to the intervention were analyzed according to predicted patterns of metaphor development, to inform discussion of the relationship of the study to other research in the field of early childhood education.
Theoretical Assumptions

Several theoretical assumptions formed the foundation for this study. These assumptions centered around three primary areas: the structure of metaphor and the process of metaphor interpretation; the developmental process of figurative competence; and the importance of cognitive mediation to encourage student learning within a zone of proximal development.

With regard to metaphor itself, the intervention utilized in the study is based on the definitional structure of a metaphor as an explicit or implicit comparison of two terms from different conceptual categories, the topic and vehicle, that are related on the basis of some shared characteristic or ground. The intent of the metaphorical structure is to impart information or emphasize features of the topic term by using salient characteristics of the vehicle to highlight less salient characteristics of the topic (Vosniadou, 1987). Moreover, the process of understanding this metaphorical structure involves a semantic mapping process, in which the individual's knowledge of the vehicle and recognition of the key ground allow said ground to be mapped to the topic in such a way that it becomes a relevant descriptor of the topic and thus enhances understanding of it (Johnson & Pascual-Leone, 1989). This theory of semantic mapping around the metaphor structure is the fundamental assumption around which key elements of the intervention are organized, with the notion that targeted instruction in metaphor definition and guided examples of mapping can improve performance on tasks of metaphor comprehension.

The second area of theoretical grounding for the study concerns the developmental process of achieving figurative competence. Although many studies reveal evidence of some behaviors related to figurative competence even in the preschool years...
(Vosniadou, 1987), there is nevertheless agreement among researchers that competence around metaphor comprehension and production are achieved developmentally, through a process that involves solidifying conceptual understanding and developing linguistic fluency in reading, writing, listening, and speaking (Johnson & Pascual-Leone, 1989; Levorato, 1993; Levorato & Cacciari, 1995). Moreover, stages in this developmental process may be defined by specific behaviors regarding the demonstrated comprehension and use of figurative language, although these stages are not entirely linear because of the influence of domain-specific knowledge levels involved in comprehension (Johnson, 1991; Levorato, 1993). However, the developmental nature of the process, the existence of measurable behaviors related to this development, and the organized structure of metaphor itself suggest that it is an area that may lend itself to cognitive mediation under Vygotsky's (1978, 1986) theory of the zone of proximal development.

Vygotsky's (1986) notion of cognitive mediation involved direct teaching of scientific or theoretically-defined concepts to students, along with the processes involved with utilizing them within a given discipline, in order to support grounding in scientific concepts as opposed to experientially-derived spontaneous concepts. He also emphasized the role of metacognitive mediation, which involved guiding students to develop strategies for planning, monitoring, and evaluating their own learning and behavior. Overall, the key element in Vygotsky's theories on instruction and learning was the importance of social interaction and language experiences between adult and child and among children (Howe, 1996). Moreover, instructional interactions between adult and child should provide children with opportunities to achieve tasks they cannot achieve independently, thus utilizing the zone of proximal development to lead children to more
advanced developmental levels. Specifically, "the only good kind of instruction is that which marches ahead of development and leads it; it must be aimed not so much at the ripe as at the ripening functions" (Vygotsky, 1986, p. 188).

In sum, then, this study assumes that figurative competence is a developmental process through which young children progress, and this developmental progression can be supported and advanced through cognitive mediation around the structure of metaphor and the semantic mapping that guides metaphor comprehension. Furthermore, the study assumes that such instruction in metaphor will be challenging and oriented within the zone of proximal development even for verbally advanced students in the primary grades because it deals with the conceptual underpinnings of metaphor rather than only on concrete examples, thus deepening the complexity (VanTassel-Baska, 1994, 1995), and because of the wide range of potential responses to metaphor tasks that can be used to determine existing levels of figurative competence (Johnson, 1991).

**Purpose and Research Questions**

The study had two primary purposes. One was to explore demonstrations of figurative competence by verbally talented second graders in the context of an intervention related to the structure of metaphor and to relate findings to predicted patterns of performance and to scores on other measures of verbal and general ability. The second purpose of the study was to determine the effects of the specific instructional intervention, designed around the definitional structure of metaphor and processes of semantic mapping, on student performance on measures of figurative competence. Thus, the following research questions were used to guide the study:
1. Do student scores on measures of figurative competence relate significantly to scores on standardized ability and achievement tests in the verbal areas and in general cognitive ability?

2. To what extent are abilities related to figurative competence demonstrated in student products completed during an instructional intervention focused on metaphor?

3. Does instruction in the definitional structure of metaphor and use of semantic mapping make a significant change in performance on measures of figurative competence assessing metaphor comprehension and literary analysis and interpretation in second grade, verbally advanced students?

4. To what extent does student performance on written measures of figurative competence relate to teacher evaluation of overall student performance on written and oral tasks completed during the intervention?

**Rationale**

One of the central goals of educational research is to provide information that can help to inform and improve school practice. This information may be provided from several perspectives on learning, development, and instructional effect. Some research provides information about how people grow and learn, thus providing a foundation for effective practice based on the needs of learners. Other research specifically investigates practices to determine their effectiveness. Information from both types of research is useful in guiding appropriate educational practice for the future.

Results of this study can provide educators and researchers with useful information in several areas of interest. First, the study can add to the existing research
base on metaphor development by demonstrating the effects of an instructional program
designed for classroom use over a period of several weeks on young children’s
performance. This would be an extension of prior studies that employed shorter
instructional periods to measure effects on performance, as well as providing data
regarding the effectiveness of direct mediation of the cognitive processes involved in
metaphor comprehension. Second, the study can support the research on a given
developmental model of figurative competence (Levorato, 1993) by relating Vygotsky’s
(1978) theory of the zone of proximal development to the abilities and behaviors
proposed in the model. Third, the study can add to the limited research base on the effects
of instructional interventions with young children of high ability and potentially
appropriate tools for differentiation for this group. In addition, because the specific
intervention was designed for classroom use and because the study was carried out within
the regular course of second grade language arts instruction, the study can also inform
recommendations for differentiation practices with young, verbally able students. Finally,
the study can provide data on the effects of a specific unit of study for young gifted
children, thus strengthening a base for further development of curriculum using the same
model.

Summary of Methodology

The study utilized a quasi-experimental, pretest-posttest comparison group design
(Campbell & Stanley, 1969) to measure effects of the intervention on a treatment group
of second grade students. Data sources included two sets of pre- and post-assessments;
student products; teacher assessment of student performance; and student scores on
standardized measures of general and verbal ability. In addition, classroom observations
were conducted in all treatment classrooms to ensure fidelity of implementation of the intervention and to collect observational records of student engagement with the assigned tasks. All instruments were piloted with relevant populations prior to their use in the study.

Data analysis techniques included analysis of covariance, correlational analyses, and multiple regression analyses of quantitative data (Creswell, 1994; Grimm & Yarnold, 1995; Kiess, 1996). Qualitative content analyses were also conducted to classify student product data according to patterns predicted by previous studies of metaphor development, according to case study methods (Yin, 1994).

Delimitations and Limitations

The study was delimited by several methodological decisions within the overall design. First, only second grade students in a single school district were used in the treatment group, rather than exploring the instructional outcomes across grade levels or across geographical sites. This constrained selection of classes was based on the challenges inherent in a classroom-based intervention study; the use of a single school district and grade level allowed for greater control over the implementation process.

Another delimitation for the study was the decision to use intact groups of students rather than random selection or assignment. The use of intact groups was, again, a direct result of the intent to find the effects of an instructional program in the classroom rather than the effects of a laboratory-based experiment. In order to implement a study in the natural setting of the classroom over a series of lessons, it was necessary to use already existing class structures.
Several limitations of the study resulted from the sampling and procedures used. The use of intact groups rather than random selection and/or assignment of individual subjects limits the generalizability of the outcomes. Additionally, the use of only five teachers for the treatment group limits generalizability, and differences among the teachers may have influenced the variable performance of their students. Issues related to the comparison group created additional limitations; response from the school district used in the study yielded only two comparison teachers, of whom one was eventually dropped from the study based on insufficient data collected, leaving only one comparison teacher within the study school district. An additional comparison class was drawn from a school in another district which uses similar methods of selecting second graders for advanced language arts work, but the comparability of the groups was weaker because of using different sites. Moreover, the comparison group remained very small even with the added class. These issues of comparability and small sample size suggest that all findings should be treated with caution.

A specific instructional unit was used as the intervention, so results may be closely related to that specific intervention rather than to instruction in metaphor in general, and the measures of performance were nonstandardized. An additional limitation of the study was that standardized test data on students' verbal or overall cognitive ability were not available prior to the beginning of the study, so students were selected for participation in the unit based on a district-developed reading performance test that classifies them at or above grade level in literacy, further limiting generalizability of the study to other populations of second graders outside the district.
Definitions of Related Terms

The following terms represent key components of the study and are used throughout as defined below. Additional discussion regarding several of the terms as they are used in research and theoretical literature is included in Chapter II.

Advanced Reader

Although the original intent of the study was to focus on an instructional intervention with young gifted students, local policies surrounding gifted identification prevented selection of an identified gifted population. In addition, local requirements surrounding reading instruction required that only those students who met certain reading standards could be involved. Thus, the sample population comprised advanced readers as defined by a local performance-based assessment in reading, with fluency of reading and comprehension on above-grade level texts representing the level of advanced reader. The significant relationship of reading fluency to comprehension and to metaphor interpretation have been addressed by Nathan and Stanovich (1991) and by Johnson and Pascual-Leone (1989), respectively.

Figurative Competence

Levorato (1993) defined figurative competence as “the acquisition of the ability to deal with figurative language” (p. 104). Dent’s (1986) definition of competence is more expansive: “the ability to detect the similarity between disparate domains and to use one domain to talk about, or understand something about, another domain” (p. 224). For purposes of this study, figurative competence is defined as the ability to demonstrate comprehension of the inherent comparisons of metaphors and other figurative language devices by identifying a metaphoric topic and vehicle and the grounds relating them, and
to demonstrate ability to develop figurative comparisons relating a topic and vehicle from disparate domains. Figurative competence is measured in the study by assessments that ask students to identify topic, vehicle, and ground of metaphors or to explain metaphoric statements, and by product evidence of demonstrated ability to write metaphors.

Figurative Language

Levorato (1993) noted that criteria to differentiate figurative from literal language have not been definitively established by psycholinguists. However, three characteristics may be used to distinguish the figurative: (1) a gap between a speaker's words and communicative intentions; (2) strongly held conventions used to establish new meanings; (3) greater dependence on context. Irony represents a clear demonstration of the first characteristic, and idiom is a representative example of the second. The degree to which the third characteristic is significant depends on the degree of conventionality of the statement (Levorato, 1993, pp. 101-102). There are several categories of figurative language, including metaphor, simile, personification, irony, idiom, and verbal humor; analogies may be classified as figurative if the topics they compare represent different conceptual domains and are thus metaphorically rather than literally related. This study will focus primarily on the metaphor, simile, and metaphorical analogy.

Gifted

The term "gifted" is generally used to refer to those individuals who demonstrate the ability to perform certain cognitive skills at a higher level than their age peers (Gagné, 1995; U.S. Department of Education, 1993). The purpose of identifying students as gifted is to target those students whose abilities differ enough from the average that they need special educational services to provide challenging learning opportunities in school.
(Feldhusen, 1998). In this study, the instructional intervention was designed with gifted students in mind, and indeed the majority of the students involved in the study were identified as gifted under their local definition during the study time frame. The local definition for gifted identification at grade 2 is based on multiple criteria, including parent and teacher recommendations, student products, and scores on several standardized instruments. Students who are recommended for screening may automatically qualify if either of the following two criteria is met: (1) Raven Progressive Matrices raw score at or above 90th percentile on national norms; or (2) at least two subtest scores on Test of Cognitive Skills in 8th or 9th stanine on national norms (Office of Instruction and Program Development, 2000). Thus, gifted in the local context is defined as scoring within the top tenth to fifteenth percentile on these nationally standardized tests, although individual consideration is then given to all students who are near to but do not meet these criteria.

High/Low Salience

Salience in metaphor study refers to the degree to which a ground characteristic is manifest in a metaphoric statement. In metaphors that have been judged to be most interesting and effective by adult and child readers, vehicle terms have high salience and topic terms low salience of the relevant ground (Ortony, Vondruska, Foss, & Jones, 1985).

Metaphor

A metaphor is a figurative language device that "communicate[s] something about a concept by comparing it or juxtaposing it to a similar concept from a different conventional category" (Vosniadou, 1987, p. 871). The word metaphor may be employed...
in both a general and a specific sense; in its general sense, it can refer to any figurative language device that makes such a comparison as explained above. In its specific sense, a metaphor is a statement that likens a “topic” to a “vehicle” on the basis of a shared feature or “ground” but does not make the comparison using terms such as *like* or *as*; a figurative comparison employing these terms would be instead labeled a simile. In this study, the general sense of metaphor is used most frequently to simplify references to figurative language that creates comparison of different conceptual categories. Use of metaphor in its specific sense, as distinguished from simile, is noted as such.

*Metaphor Comprehension and Metaphor Interpretation*

Metaphor comprehension is the process of understanding the relationship expressed between topic and vehicle in a metaphor and demonstrating that comprehension through some observable behavior. Metaphor interpretation refers specifically to the demonstration of comprehension by restating a metaphor in one’s own words (Johnson, 1991). In this study, metaphor comprehension is used as the more general term, including the behavior of selecting an appropriate translation of a metaphor from a set of choices; interpretation specifically refers to responses students must produce without a set of choices.

*Metaphor Production and Metaphor Development*

Metaphor production refers to the behavior of stating a metaphor aloud or in writing that demonstrates an understanding of different conceptual categories and the ground relating them (Vosniadou, 1987). Metaphor development, on the other hand, refers to the broad developmental process of acquiring skills related to metaphor comprehension, interpretation, and production, and in this study represents the metaphor-
related aspect of the development of figurative competence (Levorato, 1993; Vosniadou, 1987).

**Semantic Mapping**

Johnson (1991) defined the process of metaphor comprehension as a process of semantic mapping, in which “in comprehending a metaphor, a person selects some semantic aspect or facet of the vehicle’s referent…and maps it onto the topic” (p. 472). In other words, the ground (or facet) of the metaphor is defined implicitly as it relates to the vehicle and then applied to the topic in a transformed way that demonstrates understanding of how it relates specifically to the topic, instead of specifically to the vehicle. The degree of transformation manifest in an interpretation of a metaphor has been shown to be related to other aspects of metaphor and linguistic development (Johnson, 1991, 1996; Johnson & Pascual-Leone, 1989).

**Zone of Proximal Development**

According to Vygotsky (1978), the zone of proximal development is “the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers” (p. 86). This concept assumes that a child’s developmental level around a given skill may be raised through instruction. Vygotsky also argued that in addition to the existing variance between different children’s actual developmental level, there is also variance in the zones of proximal development from one child to another. Thus, although two children’s developmental levels might both be raised by scaffolding, one child may grow more than the other. In this study, the concept of zone of proximal development is related to
metaphor development in the assumption that scaffolding around the definition of metaphor and semantic mapping can support students in advancing their metaphor development.

Conclusion

The study was designed to function at several levels to investigate a specific developmental area in young children and to explore means for increasing the body of knowledge about interventions for a particular population. The role of metaphor development as a key part of overall linguistic competency is important to investigate because of the conceptual challenge provided by the interpretation of a metaphor, because of the potential to use metaphors to introduce new concepts to readers and listeners, and because of the richness that metaphor provides in literature and in speech. Focusing on metaphor development in young children encourages the use of strong literature that includes powerful imagery and invites children to explore beyond their literal interpretations into abstract ideas and connections across elements of the world around them. The intervention included in this study was designed to provide foundation in the structure and function of various forms of figurative language, to expose children to some exemplary uses of metaphor in literature, and to invite them to raise their own developmental level in terms of both comprehension and production of figurative language.

The emphasis on differentiated instruction for young gifted children is a response to the problems of serving a group of children who are not frequently identified formally yet who require special services to meet their needs for challenge and developmental appropriateness in the classroom. The unit utilized in the study, in addition to being based
on psycholinguistic research on the nature of metaphor, is also designed to challenge the advanced verbal skills that young children gifted in the verbal areas frequently demonstrate, including introducing them to advanced vocabulary and literature, discussing abstract concepts, and making use of methods of reasoning that encourage metacognitive behaviors.

With these two primary areas of focus in mind, the emphasis on metaphor development and on appropriate curriculum for young gifted children, the next chapter will focus on the nature of metaphor itself, the development of figurative competence, and the directions for research provided by past work in these areas. Moreover, the review of the literature will examine the existing research and recommendations regarding early childhood gifted education and curriculum for the gifted in general, to provide a basis for the proposed study and the intervention that it encompasses.
CHAPTER II

Review of the Related Literature

Introduction

Metaphors are a pervasive part of our everyday life and speech, in the academic realm as well as in personal and social realms (Lakoff & Johnson, 1980; Manning & Wray, 1990). Studies of speech have revealed that as many as one in eight of people’s utterances contain nonliteral references or figures of speech (Nippold, 1985); metaphors, idioms, and proverbs pervade our literature and appear over and over again in the media, in advertisements, and in personal conversations. In light of this prevalence of figurative language in speech as well as in literature, the nature of metaphor and how human beings come to understand it have long been topics of study in the areas of language, literature, psychology, and education. Indeed, attempts to define the nature of metaphor go back as far as classical times, when Aristotle discussed the meaning of metaphor and its role in knowledge and understanding of language (Addison, 1993). While literary theorists discuss and debate the nature of metaphor, developmental psychologists have over the last twenty-five years turned their attention to how people grow to understand and produce metaphors, and educators have worked to relate those findings to how language and literature are taught in schools.

This study used an exploration of the nature of metaphor and previous findings regarding metaphor development and figurative competence in general as the foundation for an investigation of the effects of an instructional intervention, focused on metaphor, within the context of a language arts program for advanced readers, many of whom were expected to be identified as intellectually gifted, in second grade. Thus the goal of the
literature review was to ground the research on metaphor development within several other frames of reference, namely (a) the larger domain of literacy instruction in the primary grades, (b) the special challenges faced by highly able students in the primary grades, and (c) the efforts within the field of gifted education to identify and serve the needs of highly able students through curriculum differentiation and instructional interventions. Thus, the literature review is organized around the following four strands:

1. Overview of theory and research regarding early literacy development, with exploration of the trend of reading instruction toward an emergent literacy framework;

2. The study of metaphor and the development of metaphoric or figurative competence as a part of linguistic and literacy development, with a review of research and theory on the developmental stages of figurative competence in children;

3. Research on ability-related differences in figurative competence in children, with particular emphasis on young gifted children and their specific characteristics and needs; and

4. Research on the role of instruction in the development of figurative competence, research on effective curriculum and instruction for highly able students, and discussion of the potential interplay of the two in supporting metaphor development within a language arts curriculum framework related to cognitive and emergent literacy theory.
Strand I: Early Literacy Development and Instructional Trends in Reading

Reading and writing instruction have long served as the centerpiece of the early elementary curriculum, with the goal of helping students learn the formal rules of written language while continuing a process of oral language development that began in infancy (Strickland & Feeley, 1991). Because of the prominence of language arts in the primary curriculum and the resultant quantity of research and writing on the subject, considerable debate exists regarding approaches to teaching reading and writing, with extensive research on how children develop literacy as the background for the discussion. Tierney (1991) noted that “understanding how literacy develops is a prerequisite to responding to readers and writers and to planning their educational experiences” (p. 176). Thus, notions of literacy development and common school practices related to them are important to examine in any study of elementary language arts.

Emergent Literacy and the Reading Wars

Over the past two decades, much of the developmental literacy research has focused on the notion of “emergent literacy” in studying young children’s reading and writing development, defining emergent literacy as the reading and writing behaviors of young children that precede and develop into conventional literacy (Sulzby, 1989, 1991; Whitehurst & Lonigan, 1998). This concept of emergent literacy places the child as an active participant in the literacy development process, arriving at school already negotiating among language stimuli to create, connect, and modify his or her conceptual structures and to develop facility with language production in speech and in writing (Tierney, 1991). Emergent literacy, then, relates to the overall process of language acquisition children undergo in their early years, testing hypotheses and generalizations.
about language through their own linguistic interactions with others, and through this process evolving their metalinguistic awareness or mature understanding of language concepts (Sulzby, 1991).

Within this framework of emergent literacy, two distinct skill areas for development and instruction may be identified: technical skills such as phonological and alphabetic knowledge, and more conceptual skills of comprehension and broader knowledge of language (Whitehurst & Lonigan, 1998). A central question for educators, then, is to what degree children may develop the technical skills implicitly as they experience spoken and written language and to what degree these technical skills require direct teaching. The fundamental assumptions about literacy learning that underlie responses to this question are the basis for the so-called “great debate” and or “early reading wars,” with opposition between approaches that emphasize code development versus meaning development (Calfee & Norman, 1998; Taylor, Anderson, Au, & Raphael, 2000). This opposition has been realized in recent years in arguments over direct phonics instruction versus literature-based whole language approaches.

Proponents of phonics instruction argue that direct teaching of foundational concepts such as phonemic awareness and letter-sound relationships will have a greater effect on student reading achievement than approaches that teach phonics more indirectly, especially among children from low socioeconomic backgrounds (Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998). One admittedly oversimplified description of this philosophy suggests that reading is itself “a basic skill – the translation of print into some equivalent of speech – to be taught by direct instruction [before getting to] the good stuff” (Calfee & Norman, 1998, p. 244). On the other side of the debate,
supporters of the whole-language approach cite research demonstrating positive effects of holistic, literature-based reading instruction that teaches coding skills indirectly (Daniels et al., 1999), with an emphasis instead on reading as "a natural extension of spoken language; immersing young children in quality literature leads naturally to competence and interest in handling print" (Calfee & Norman, 1998, p. 244).

Research reviews mediating the debate have demonstrated positive effects across the range of approaches, concluding that a balance between code instruction and meaning instruction is the key to supporting literacy development (Au, 1998; National Reading Panel, 2000; Snow, Burns, & Griffin, 1998; Taylor et al., 2000). These recommendations also emphasize the importance both of authentic literary activities around many genres of literature and of explicit skill teaching, although arguing that the latter should not be the dominant component in a language arts program (National Reading Panel, 2000). The balanced perspective may be summarized as an approach to literacy that supports the "acquisition of a linguistic register, of strategies for using language in a planned and thoughtful manner" (Calfee & Norman, 1998, p. 244). This approach also places a strong emphasis on social interaction between teacher and learner and among learners as a key part of the process, reflecting Vygotsky's (1986) theory that social interaction and language are themselves the basis for the development of thought, conceptual understanding, and the metacognitive processes that govern development in learning and behavior. The relevance of Vygotskian theory in literacy development and in metaphor development in particular will be discussed further later in this chapter.

In addition to emphasizing balance in literacy instruction, researchers also suggest that the complexity and individual variability of the literacy development process in
different children limits generalization and prevents the identification of one approach as a “silver bullet” to solve the puzzle of how all children best learn to read (Foorman, Fletcher, Francis, & Schatschneider, 2000; Taylor et al., 2000). This individual variability among children in their literacy development relates to many factors, including exposure to environmental print, types of linguistic interactions with adults, and measurable verbal ability differences (Jackson & Klein, 1997; Sulzby, 1991). The complexity and variability among children and the differential research findings around various approaches support the notion of the balanced approach to literacy instruction. Moreover, they support the theoretical notions around developmentally appropriate practice (Bredekamp & Copple, 1997), especially the recognition that students both begin at different levels and advance at different rates within any given domain of development.

**Precocious Readers**

An example of this variability may be shown in the research on precocious readers. Precocious readers are those children who are able to read and comprehend texts prior to school entry, beyond the ability merely to repeat a familiar book from memory with page-turns at correct moments (Jackson & Klein, 1997; Sulzby, 1991). Jackson (1988, 1992) noted that many precocious readers rely heavily on a large sight word vocabulary before figuring out the phonological system, so they often begin school able to read many texts but without strong phonological decoding skills. However, although this pattern and a tendency to be able to read text very quickly are common among precocious readers, they nevertheless also demonstrate great differences in the relative strengths of their various linguistic skills (Jackson & Klein, 1997), and precocious reading is not always an indication of an advanced verbal IQ (Jackson, Donaldson, &
Cleland, 1988). Moreover, precocious reading does not alone constitute advanced literacy under emergent literacy definitions, because it does not necessarily encompass advanced writing skills. Nevertheless, within the population of precocious readers are some who do develop phonological decoding skills earlier and more quickly than their age peers, placing them in the difficult position of already knowing much of what will occur in a skills-based language arts program in the early grades.

This discussion of precocious reading ability demonstrates one of the central difficulties involved with language arts instruction in the primary grades, namely that children vary significantly in the knowledge and experiences that they bring to the language arts classroom and in their stage of literacy development (Foorman et al., 1998). The consequent complexity in planning instruction that will be developmentally appropriate for such a range of students is evident in the continued debate regarding time spent teaching phonological skills - even to those children who are reading numerous sight words but not necessarily decoding - versus time spent engaging children in literature-based language experiences to support implicit development of technical reading skills.

*Early Literacy and Metaphor*

Within this broad debate over philosophical positions regarding reading instruction, where does discussion of metaphor belong? The pervasive presence of figurative language within literature selections suggests that children engaged in a whole-language program or one balanced between literature study and skill development will have opportunities to experience metaphor as they read (Winner, 1988). The potential power of metaphor instruction in emergent and early literacy is in its requirement that the
reader or listener make meaning of language. Strickland and Feeley (1991) noted that "meaning should be at the center of all reading activities" (p. 294) and that a language arts program should focus on broadening a child's conceptual and experiential base, with the vocabulary for reading and writing growing as a consequence. Writing instruction around metaphor gives students opportunities to put new concepts and vocabulary to use and to experiment with the conventions of language, one of the major processes involved with emergent literacy (Sulzby, 1991).

Moreover, the purpose of a metaphor is to draw the reader or listener's attention to a new conception of a given topic, to extend an understanding of meaning by connecting one idea to something with shared characteristics in a different conceptual category. Consequently, metaphors can expose children to new words and concepts even as it supports them in advancing understanding of what they already know. In addition, metaphors vary significantly in level of complexity depending on the conceptual categories they compare and the grounds for these comparisons (Vosniadou, 1987). Thus, study of metaphor can offer challenge to students functioning at different levels of linguistic development and with different degrees of background knowledge, making it a strong source for providing differentiated experiences for young students.

To establish a foundation for this role of metaphor in language arts instruction, a discussion of the psycholinguistic theory around metaphor and the research on development of metaphoric competence is warranted.
Strand II: The Nature of Metaphor and the Development of Figurative Competence

Theories of Metaphor

According to Vosniadou (1987), "metaphors are meaningful statements that communicate something about a concept by comparing it or juxtaposing it to a similar concept from a different conventional category" (p. 871). The basic description or structure of a metaphor is that it contains a term (the "topic") which is likened to another term (the "vehicle") on the basis of one or more shared features (the "ground"). However, although metaphors can thus be defined and described relatively simply, a deeper understanding of their nature has been the subject of much debate. Three main views exist on the nature of metaphor: (1) the substitution theory, which suggests that a metaphor merely involves the replacement of a literal term with a figurative one; (2) the comparison theory, which suggests that a metaphor asserts a similarity or comparison between the topic and vehicle terms; and (3) the interaction theory, which suggests that a metaphor demonstrates a relationship that allows the topic to be seen from the perspective of one's knowledge about the vehicle (Vosniadou, 1987; Winner, 1988).

Winner (1988) discussed each of these theories and their relative merits, concluding that the substitution theory is the weakest, in that it assumes a parallelism between metaphoric statements and literal translations of them; this parallelism may not always exist, nor can every metaphor be translated into a literal statement. Furthermore, the substitution theory pays little attention to the issue of fundamental similarity between the topic and vehicle in a metaphor. Although she found the comparison theory to be stronger, in that it acknowledges this fundamental similarity, Winner also suggested that the comparison theory is limiting because it equates the comparison process used to
understand metaphor with its actual nature. Furthermore, the comparison theory does not clearly distinguish between the separate functions of the topic and vehicle in the metaphoric statement.

The third theory, the interaction theory, remains as the strongest of the three, in that it does not assume a literal substitution, nor does it limit a metaphor’s nature to a basic comparison. Rather, it asserts the significance of the relationship between topic and vehicle and suggests that the vehicle (as its name metaphorically suggests) takes one to a deeper understanding of the topic through the nature of the relationship. Johnson and Pascual-Leone (1989) described the process involved with understanding metaphor as semantic mapping, in which first the vehicle and its referent ground are recognized, then the key features of that ground are mapped, or applied, to the topic to demonstrate a key feature of that topic. They also noted that developmental differences in metaphor comprehension are evident depending on the degree to which the ground is transformed in interpretation to apply specifically to the topic.

As this semantic mapping view indicates, the interaction theory requires a fundamental assumption of asymmetry in a metaphor. The vehicle is selected precisely because of what it can demonstrate about the topic, and the two cannot easily change places because of their relative degrees of salience with regard to the ground. In a literal comparison, both terms are generally high in salience around intersecting attributes. In a figurative comparison, on the other hand, generally only the second term is high in salience, because it is the strength of the ground in this vehicle that encourages the new understanding of that characteristic’s importance in the topic (Glucksberg & Keysar, 1993). Indeed, studies have been conducted demonstrating that participants judged
figurative comparisons to be most effective when the second term was high in salience and the first was low with regard to the ground characteristic (Ortony, Vondruska, Foss, & Jones, 1985).

**Literal Versus Figurative Language**

Another significant aspect of understanding the nature of metaphor is the issue of drawing from different conventional categories. Studies of metaphor generally agree that statements relating objects within the same category are not metaphorical but rather literal comparisons (Vosniadou, 1987, 1995); asserting that a lion is like a tiger or that a pine and a fir share some characteristics does not require the level of figurative interpretation as a comparison of a soldier to a lion or a tree to a statue. However, conventional categories are rarely distinct and absolute; their boundaries may be difficult to define: “the figurative is associated with definitive difference; but difference manifests itself to various degrees and in various ways...” (Addison, 1993, p. 416). Thus, a certain amount of judgment may be necessary in determining whether an expression is metaphorical. Indeed, although a metaphor by definition cannot be an entirely literal statement, metaphors exist along a continuum from more literal to more metaphorical, depending on the remoteness of the categories of the topic and vehicle. The more remote the categories, the more metaphorical the statement (Vosniadou, 1987).

Although the distinction between literal and metaphorical statements cannot be clearly distinguished in all cases, three important characteristics of figurative language help in determining the difference and in understanding the process of acquiring figurative understanding (Levorato, 1993). The first relates to the issue of speaker intent. The speaker or writer of a metaphorical statement recognizes the existing gap between
the literal words and the communicative intention, or, in other words, recognizes a violation of conventional categories occurring in the statement, yet links the topic and vehicle deliberately (Levorato, 1993; Vosniadou, 1987). Indeed, the greater the dissonance between stated terms and implied meaning, the more figurative a statement may be judged to be (Winner & Gardner, 1993). This issue of intent has played an important role in research addressing the development of metaphoric competence, as will be discussed below.

The second characteristic of metaphor is its difference from literal statements in terms of the need for judgments to be made: “a comparable literal statement may be judged in terms of its truth, whereas a metaphor engenders a judgment about a new meaning” (Palermo, 1986, p. 15). Thus, comprehension of a metaphor implies not only recognition of a relationship, but also a judgment as to whether the metaphor informs understanding of the topic through that relationship. Finally, figurative language is generally more dependent on its context than literal language for understanding (Levorato, 1993); as with the issue of intent, the contextual basis of metaphor has played an important role in research on metaphor development in children.

Types of Metaphors

Vosniadou (1987) identified six types of conceptual relations representing possible areas of connection between topics and vehicles: (1) descriptive properties, (2) characteristic activities, (3) emotions and thoughts, (4) structural/functional characteristics, (5) causal properties, and (6) plans and goals. These conceptual categories represent sources for the grounds linking topic and vehicle. In demonstrating these connections, there are two main types of metaphors: predicative and proportional.
Predicative metaphors represent a one-to-one relationship between topic and vehicle; generally, the topic and vehicle are joined on the grounds of some attribute clearly distinguishable in both. Proportional metaphors, on the other hand, represent a two-to-two relationship, though very often one of the topics is unstated. The grounds of proportional metaphors are usually relational rather than attributional. Thus, if stated more explicitly, they would be analogical statements that demonstrate that the first topic is related to its vehicle in the same way that the second topic is related to its respective vehicle (Manning & Wray, 1990).

Gentner (1988) used a similar categorization structure in discussing types of metaphors, naming those based on mere appearance characteristics attributional metaphors, while those based on functional characteristics were called relational metaphors. Further, she noted that some metaphors represent a combination of attributional and relational, calling such comparisons double metaphors. Gentner also discussed reasoning by analogy as a central element involved in the process of interpreting relational and double metaphors.

The connection between metaphor and analogical reasoning is an important one in terms of understanding how topic and vehicle are related and how this relationship is understood (Castillo, 1998). An analogy may be used to express all the elements of a proportional (or relational) metaphor, demonstrating clearly how topics and vehicles are linked to one another. Vosniadou (1995) defined analogical reasoning as “the identification of the correspondences between two systems and the transfer of relational information from one system to the other” (p. 300). She noted, however, that metaphor cannot be drawn from within-domain analogies – those analogies that demonstrate...
relationships within items in the same category – because this would not satisfy the
requirement that a metaphor must contain a topic and vehicle drawn from different
conventional categories, else it becomes only a literal comparison.

The discussion thus far has focused on general attributes of metaphor as a broad
representation of figurative language. However, within the general category are many
different types of figurative language, each of which has singular characteristics. One of
the most common of these types is the simile. Metaphor in the specific sense refers to a
figurative comparison of topic to vehicle on the basis of some ground that usually is not stated; a simile, on the other hand, makes the comparison more explicit by adding “like” or “as” or similar words, and frequently also by stating the ground on which the topic and vehicle are being related. Although simile and metaphor are frequently classed as variations on the same theme, the apparently surface level distinction between them actually masks more complex differences. Addison (1993) discussed the debate over whether metaphor and simile should be connected in the first place. The continuum between literal and figurative becomes even wider in considering similes, because similes can lean further toward the literal than metaphors can (Glucksberg & Keysar, 1993). Indeed, “a simile, though it cannot actually express identity or opposites, can express any among an infinity of degrees of likeness and unlikeness” (Addison, 1993, p. 404). Another issue debated among students of figurative language is whether the two terms of a simile must be connected levelly, using “like” or “as,” or whether statements based on a relationship of “like but greater than,” etc., can be classed as similes. Many cite Milton’s work as an example of masterful use of comparisons of this type (Addison, 1993).
Other types of figurative language may also be identified, including personification, which grants human characteristics to non-human topics; irony, through which the speaker's intent is the exact or near opposite of the words spoken; and verbal humor, which uses the ambiguity created by the use of a term with more than one meaning as the source for humor (Manning & Wray, 1990). These other categories, however, are less the subject of focus in the research because frequently they may be classed under simile or metaphor. Furthermore, the proposed study, although it will employ personification and verbal humor in the course of the instructional methodology, focuses more specifically on metaphor, simile, and the analogical reasoning that supports them than on these other types of figurative language, in an attempt to explore the process a child engages in to develop an overall figurative competence.

*Figurative Competence*

Figurative or metaphorical competence is "the ability to detect the similarity between disparate domains and to use one domain to talk about, or understand something about, another domain" (Dent, 1986, p. 224). Thus, it requires a certain amount of cognitive sophistication, as well as linguistic skills. Levorato (1993) claimed that the development of figurative competence is inextricably linked to the development of linguistic skills, and that figurative competence does not encompass a separate set of skills from literal competence. According to Levorato, "the acquisition of figurative competence is tied to the development of a whole series of linguistic skills that give the child an ever greater control over his or her communicative possibilities" (p. 119). Among these important linguistic skills are abilities in coding, making inferences, activating world knowledge, using imagination and creativity, and finding out the
communicative intent of the speaker. The culmination of the development of these varied skills and other related ones, according to Levorato, is a level of metalinguistic competence that involves careful reflection on language itself. Similarly, Winner and Gardner (1993) emphasized that beyond the ability to comprehend metaphor is an advanced level of metalinguistic awareness; at this level, consciousness of the disparity between what is said and what is meant in a nonliteral utterance allows a listener or reader to appreciate the metaphorical nature of the statement.

Leverato (1993) suggested the existence of a clear link between development of literal competence and of figurative competence based on a series of reasons tied to both economy of cognitive functioning and the nature and use of figurative language itself. She argued that because of the prevalence of figurative language in everyday speech and writing and because of the inefficiency implied by the idea of different processing for two varieties of linguistic stimuli, the comprehension and production of figurative language cannot be that different from that of literal language. Such a notion of economy of cognitive processing relates to other research assessing to what degree cognitive skills are context-bound and the relationship between general strategic knowledge and domain-specific knowledge (Chi, Hutchinson, & Robin, 1989; Perkins & Salomon, 1989). These researchers have argued that general strategic knowledge and domain-specific knowledge interact in the learning process, with certain skills spanning content domains and transferring from prior knowledge to new knowledge, while others require specific focus and development within domains. Within the linguistic context, this perspective reflects Leverato’s (1993) understanding that the same generalized linguistic processing skills are at work with literal and figurative language development, although some specific skills
and knowledge must be developed related primarily to expressions in the figurative. These specific skills include the ability not only to comprehend the figurative statement, "but also [to mark] the utterance as a special form of speech," once again emphasizing metalinguistic awareness (Winner & Gardner, 1993, p. 427).

Thus, literal and figurative language may be considered ends of a continuum with a range of linguistic constructions in between. Consequently, the development of figurative and literal competence in understanding and producing language are closely linked and use many of the same processes, suggesting that attention to children's development across the continuum of competencies can be beneficial to overall linguistic ability, and thus, that study of figurative language within the language arts educational program is important for developing a wide range of linguistic skills. The result of development in both of these areas is to reach a level of competence in searching for the "greatest possible degree of coherence among all the linguistic and nonlinguistic information processed at a given moment" (Levorato, 1993, p. 104).

Also entering the complex configuration of linguistic development related to literal and figurative language is the added variable of content knowledge related to the concepts expressed in a given metaphor. Familiarity with vocabulary and with the specific content areas referenced by the topic and vehicle of a given metaphor also may affect the degree to which a child is able to demonstrate understanding of that metaphor. Winner and Gardner (1993) argued that limited domain-specific knowledge was the primary if not the only barrier to metaphor comprehension in young children. Johnson (1991), however, argued that metaphor development is affected by three main areas: general cognitive capacity, overall linguistic ability, and domain-specific knowledge.
Thus the developmental process related to figurative competence is complex and not linear or one-dimensional; rather, it can demonstrate itself at different levels of development depending on a number of other variables.

Model of Figurative Competence

With the clear links between figurative and literal competence in mind, and with the assumption that other variables of general cognitive capacity and domain-specific knowledge may affect level of competence in interpreting figurative expressions, attention must then turn to those specific abilities that support figurative competence within a larger metalinguistic awareness. Figurative competence may be distinguished as “a coordinated set of abilities, integrated within the general cognitive mechanism underlying semantic competence and language comprehension” and encompassing abilities to recognize different domains, to go beyond a purely literal-referential strategy in comprehension, to use contextual information, and to recognize that strongly held conventions of conceptual categories may be deliberately violated in the use of language (Levorato & Cacciari, 1995, pp. 263-264).

Levorato and colleagues (Levorato, 1993; Levorato & Cacciari, 1995) incorporated these specific abilities related to figurative language comprehension and production into a Model of Figurative Competence (see Figure 1), based on studies of children’s comprehension and completion of figurative language tasks. This model is based on six abilities determined to represent aspects of figurative competence:

1. the ability to gain a gradually broadening sense of word meaning, its position in a given semantic domain, and its paradigmatic and syntagmatic relations
Figure 1. The Model of Figurative Competence (Levorato, 1993; Levorato & Cacciari, 1995). The graphic lists the levels and abilities as described by Levorato; the links between the abilities and the levels is this researcher's interpretation demonstrating at which levels the specific abilities generally begin to appear.
2. the ability to understand the dominant, peripheral, and polysemous meanings of a word, and also the ability to perceive the relationship between a given meaning and other related meanings

3. the ability to suspend a purely referential strategy

4. the ability to understand the figurative uses of a word and the relationship between the literal meaning and a figurative meaning

5. the ability to process large amounts of language at once in order to identify the meaning of ambiguous or unknown expressions

6. the ability to use figurative language productively in the creation of new figures of speech by means of the lexical and syntactic transformation of preexisting figures of speech (Levorato, 1993, p. 104).

These six abilities are developed over a series of levels (Levorato, 1993). At Level 0, children identify objects totally with their respective names, to the extent that although they may recognize that an object's name may change, they believe that the properties of the object will change with the changing name (Osherson & Markman, 1985, in Levorato, 1993). At Level 1, the name of an object begins to become a symbolic substitute in the mind of a child. The child begins to understand that the same linguistic label can be given to various referents, and that different names may refer to the same object. This is an indication of a growing system of conceptualization and categorization, providing the foundation for the eventual recognition of category distinction necessary in the understanding of metaphor. At Level 2, a child moves beyond a purely referential and literal strategy, allowing context and inference to affect their understanding of meaning.
rather than relying strictly on the literal. This recognition of the role of context is also
critical in beginning to understand metaphor. Level 2 is the first level for which Levorato
(1993) provided an approximate age of the child, suggesting that children generally arrive
at Level 2 around the age of 7 or 8. At Level 3, children are able to use their contextual
and inferential strategies to recognize when the literal is infringed or when conventional
categories are violated. Furthermore, at this level children begin to realize the relevance
of the speaker’s communicative intent, as well as to understand that some ways of
expressing an idea may be more effective than others, given a context. At Level 4,
children acquire understanding and facility with figurative expressions as units,
recognizing the contexts in which it is most appropriate to use such linguistic units. The
final level of development of figurative competence, Level 5, represents the ability to
analyze and reflect upon figurative expressions, with more careful attention to the
relationship between the expression and the intent of the speaker; this level also
represents a more advanced level of production and usage of figurative language.
Levorato calls this advanced level “metalinguistic competence” (p. 119) and suggests that
the process of its development is significant in giving a child control over communicative
possibilities.

Again, application of these levels to the study of figurative competence in
children must take into account other developmental, cognitive, and experiential
variables. For example, Levorato herself noted that Level 5, especially when applied to
developmental understanding of idiom, is heavily influenced by familiarity (Levorato,
1993), although the context of an utterance was found more significant than familiarity at
native English speakers and bilingual Spanish-English speakers also found that language-related domain knowledge influenced level of metaphor comprehension, although language proficiency did not, and age was the most important variable across both groups.

Other research has also supported the notion that figurative competence is a developmental process, with children’s ability to interpret metaphors at a figurative level increasing with age (Evans & Gamble, 1988; Manning & Wray, 1990; Palermo, 1986). Such findings also relate to language acquisition and emergent literacy research that explores children’s increasing exposure to and experimentation with conventions of speech (Sulzby, 1991; Tierney, 1991). Moreover, developmental research on figurative competence emphasizes that understanding of metaphor and production of metaphor are integrally related skills, as are reading and writing under an emergent literacy framework (Tierney, 1991), but that comprehension develops ahead of production and represents a different though not entirely separate competency. Consequently, researchers have investigated both of these skills in young children to determine the contexts and stimuli which can encourage development of each.

**Research Findings on Figurative Competence**

Across the range of figurative competence studies, findings have emerged regarding the role of context and complexity of input, as discussed by Vosniadou (1987), and around Johnson’s (1991) key factors of general cognitive capacity, linguistic ability, and domain-specific knowledge. Furthermore, the studies have most frequently emphasized the role of age-related developmental differences in performance on tasks
related to figurative competence. These central issues in the study of metaphor development are addressed with a discussion of relevant studies.

Levorato’s (1993) Model of Figurative Competence demonstrates a process of development that encompasses both comprehension of metaphors and their production. Nevertheless, it is worthwhile to explore comprehension and production briefly as separate competencies, with consideration of the research on how children develop each.

Comprehension.

Vosniadou (1987), in reviewing literature demonstrating young children’s difficulty in paraphrasing or explaining metaphors, argued that this was not a result of an inability to comprehend metaphors. Rather, she postulated, several other factors could be identified as sources for failure in comprehension. First, children may not realize that a given utterance is intended to be interpreted metaphorically instead of literally. Second, children may fail to see the similarity or ground that is the basis for the metaphor. Third, children may understand the metaphor but be unable to produce an appropriate response to paraphrase or explain it. With these three factors in mind, Vosniadou identified two critical variables affecting children’s metaphor comprehension, each of which has been explored in various research efforts around children’s figurative competence, and also discussed the issue of the type of metaphor task employed in testing.

The first critical variable is context; the context of the metaphor can provide information as to its meaning as well as an indication that the utterance should be interpreted metaphorically instead of literally. Indeed, Vosniadou argued, “It may be the case that metaphor comprehension is originally achieved only in situations where the already established context strongly leads to inferences that are inconsistent with a literal
interpretation and consistent with the metaphor's implied meaning" (1987, p. 878).

Levorato (1993), as explained above, identified the relevance of context in children's development of figurative competence, particularly at Levels 2 and 3. In a study of children's comprehension of idioms, Levorato and Cacciari (1992) found that familiarity played only a minor role in comprehension and only in those children not yet able to use contextual information to derive meaning. Vosniadou and colleagues (Vosniadou, Ortony, Reynolds, & Willson, 1984) and Broderick (1991), in studying young children's comprehension of metaphor and simile situated within a story context, found children even as young as 5 or 6 able to select appropriate paraphrases of the metaphors presented.

Vosniadou's (1987) second critical variable affecting children's metaphor comprehension is the complexity of the linguistic input itself. She suggested that some types of figurative language may be easier to understand than others (e.g., a simile may be easier than a metaphor) and that some types lend themselves more definitively to a nonliteral interpretation. Nippold and Haq (1996) found that the concreteness of the statement had a significant effect on comprehension of proverbs in students at grades 5, 8, and 11. Gibbs (1991) found that children had most success comprehending idioms that were decomposable – those in which the meanings of the individual parts contributed to the overall meaning. In addition, the content of the comparison is significant; children must have a solid understanding of the words and concepts involved in the comparison, particularly of the vehicle being used and its relationship to the ground of the statement. Furthermore, as discussed previously, the relationship between topic and vehicle in a given metaphorical statement varies in complexity depending on whether the metaphor is attributional or relational, predicative or proportional, and children have been found to
more readily comprehend metaphors based on attributes of the topic and vehicle than those that are relational in nature (Gentner, 1988; Vosniadou, 1987).

Beyond these critical variables in overall comprehension, Vosniadou (1987) also noted the importance of considering the type of measure used in studying children’s figurative competence. Based on findings from a series of studies she and colleagues conducted, Vosniadou asserted that “children demonstrate a greater understanding of metaphor in multiple-choice or enactment tasks than in paraphrase tasks, presumably because the former impose fewer linguistic and metacognitive demands than the latter” (1987, p. 877). Winner (1988) also posited that “children understand metaphors before they can successfully explain their understanding” (p. 45). Broderick (1991) explored children’s performance on a series of simile tasks of varying complexity, including selection, evaluation, explication, and production, finding that competence in selection tasks appeared first (in the youngest children), followed by explication and evaluation with no significant difference between the two in terms of age of the child, and finally production appeared as the most complex of the tasks.

While recognizing that the linguistic demands of asking children to supply metaphor interpretations are greater than those of asking for selection of an appropriate response, other researchers have specifically investigated such interpretations; the focus of such studies has not been correctness of interpretation, but rather the relative complexity of types of responses children are able to supply (Johnson, 1991; Johnson & Pascual-Leone, 1989). Again with the hypothesis that metaphor interpretation occurs developmentally, these researchers asked children aged 6-12 and 20-year-old adults to interpret given metaphors and then classified their responses at different levels depending
on the demonstration provided of the topic-vehicle-ground connection. Findings demonstrated that by about age 7 or 8, children reliably produced interpretations that reflected application of the vehicle’s ground to the topic, but with no indication of the subtle differences in the way the ground related to the two terms. These were classified as Identity interpretations. At the next level, the Analogy level, reached reliably by about age 9 or 10, children were able not only to recognize the correct relationship but also to transform the ground in such a way as to indicate specific application to the topic. In other words, the children used the vehicle’s high salience attribute (or “facet”) to come to understand what important attribute of the topic was being addressed, but used the topic’s specific attribute instead of the vehicle’s to demonstrate understanding. Rather than applying some facet of the vehicle related to the topic only figuratively, the child identified the relevant facet of the topic that the corresponding ground brings to mind.

Additional levels of interpretation under this framework included two Predicate levels, at which the child elaborated on the relevant facet to demonstrate its application to the topic in specific instances (concrete experiential predicate) or in a more generalized sense (general conceptual predicate). The predicate level was reached reliably by age 11 or 12.

Beyond the age-related developmental differences evident in these studies, the researchers also investigated differences in performance between native English speakers and native Spanish speakers who had learned English as a second language (Johnson, 1991, 1996; Johnson & Pascual-Leone, 1989). Once again, the findings demonstrated that age differences had the strongest effect, and that degree of oral language proficiency in a second language did not transfer to differences in metaphor interpretation. From this
finding, the authors generalized that not only were developmental age-related differences most important in differences in metaphor interpretation, but also that understanding the underlying linguistic structure of metaphor was more important in interpretation than language proficiency.

Production.

The studies discussed to this point have primarily focused on metaphor comprehension, with a few exceptions; the discussion turns now to the related task of metaphor production. This area of research is in some ways more complex than that of metaphor comprehension and interpretation, because of the issue of assessing speaker intent (Vosniadou, 1987). Research as to the developmental progression of metaphor production in young children has demonstrated that even preschool children in play and language often display actions and utterances that at face value may be considered metaphorical (Gardner & Winner, 1978; Winner, 1988). However, a question arises as to whether these "child metaphors" may be considered real metaphors or not. Vosniadou (1987) argued that according to the definition of a metaphor, a statement is only metaphorical if the speaker intentionally violates conventional categories in making the figurative comparison. Consequently, if a young child has not yet developed solid conceptual categories, this intentional violation may not be taking place. The issue of solidifying conceptual categories is one that has been a central element of developmental research and theory across the history of cognitive psychology. Piaget (1952) and Vygotsky (1986) both emphasized that young children form their own conceptual categories about the world based on their experiences, and that these categories vary in the degree to which they reflect the conceptual categories accepted in the adult world. In
Vygotsky’s exploration of the types and origins of conceptual categories by children, he distinguished between spontaneous or everyday concepts, which they acquire through experience, and scientific concepts, which are acquired from direct teaching (Howe, 1996; Vygotsky, 1986). Thus, a young child’s “metaphorical” statement that violates conceptual categories may or may not be intentional, depending on to what degree those categories are clearly differentiated in the mind of the child.

Moreover, Gardner and Winner (1978) proposed that although preschool children are often prolific in their use of these “child metaphors,” they then go through a decline in spontaneous metaphor production in the early elementary years as they consolidate their literal understanding of words before returning to use of the figurative in preadolescence. This pattern places the suspension of the literal at a somewhat higher age level than Levorato’s (1993) model does, but both recognize children’s need to have a growing familiarity with the literal before they are able or willing to suspend the literal in favor of a figurative interpretation of words.

Nevertheless, despite arguments that children in the preschool and primary years cannot consistently produce intentional metaphors, several studies and additional anecdotal evidence have demonstrated the capacity for focused metaphor production in children in the elementary years. Rummel and Dykstra (1983) analyzed free writing samples from second, fourth, and sixth graders for spontaneous production of analogy, noting number of analogies produced and type of analogy, including personal, direct, symbolic, and fantasy. From this analysis, the researchers found that analogy production did increase with age, but only if fluency was not factored out, and that the use of symbolic analogies, which are generally the most strictly metaphorical, was not
significantly related to age. When the samples were analyzed with fluency levels accounted for, the second graders actually used analogies more frequently than the fourth and sixth. The authors concluded that children use metaphor as a writing tool actually less frequently as they move through these years, but that their use becomes more conscious and targeted, which also reflects Gardner and Winner’s (1978) notion that some metaphors from younger children may not have metaphorical intent.

Additional evidence of young children’s metaphor production was detailed by Steinbergh (1999), who offered sample writing from students in grades 1-6 and demonstrated clearly intentional metaphorical comparisons written by first and second graders. Her findings, in addition, demonstrated that the types of relationships offered in metaphors change with age, with younger children more frequently using attributional metaphors demonstrating clear sensory relationships. This use of attributional metaphors earlier than relational metaphors reflects the findings that young children also comprehend the attributional earlier than the relational (Gentner, 1988; Vosniadou, 1987; Winner & Gardner, 1993). Nevertheless, as Winner (1988) noted, researchers have found “instances of metaphor conveying abstract ideas where children are intuitively aware of the meaning and power of their image, years before they can explain how the language is working” (p. 325).

The studies outlined to this point have focused largely on age-related developmental differences in figurative competence. However, developmental differences among children are not related solely to age, because different children develop at different rates across all the domains of human growth (Bredekamp & Copple, 1997; Silverman, 1997). A few studies have examined ability differences in relationship...
to figurative competence, generally finding some aspect of ability difference to be correlated, whether general ability, specific verbal ability, or mental age (e.g., Fung, 1995; Johnson & Pascual-Leone, 1989). Because of the focus of this study on an instructional intervention with potentially gifted children, the next strand will examine those studies demonstrating ability differences and particular relationships of the findings to the literature on young gifted children.

**Strand III: Young Gifted Children and Figurative Competence**

*Individual Differences and Figurative Competence*

Studies of figurative competence as it relates to cognitive differences in children have shown varied results, again suggesting that many factors affect performance on measures of metaphor. Nevertheless, these studies generally support the notion that figurative competence, as measured by metaphor interpretation or related tasks, is related to other aspects of ability and development, as measured by standardized aptitude tests and other measures. Relevant studies are discussed below, and Table 1 summarizes the variables for which relationships with figurative competence have been demonstrated, grouping the studies based on whether subgroups within the samples were compared or whether relationships were assessed across the sample.
Table 1

*Summary of Studies Relating Ability Measures to Figurative Competence*

<table>
<thead>
<tr>
<th>Studies</th>
<th>Variable(s) demonstrating significant positive relationships with performance on measures of figurative competence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studies correlating continuous variables across sample</td>
<td></td>
</tr>
<tr>
<td>Johnson, 1991</td>
<td>Developmental linguistic ability</td>
</tr>
<tr>
<td>Johnson &amp; Pascual-Leone, 1989</td>
<td>Nonverbal measure of mental capacity</td>
</tr>
<tr>
<td>Fung, 1995</td>
<td>IQ (significant at grade three but not kindergarten)</td>
</tr>
<tr>
<td>Studies comparing performance of participants in different groups</td>
<td></td>
</tr>
<tr>
<td>Johnson, 1996</td>
<td>Second language learners of English versus first language learners of English, favoring the latter; differences significant only for subgroups</td>
</tr>
<tr>
<td>Jones &amp; Stone, 1989</td>
<td>Learning disabled versus normally functioning adolescents, favoring the latter</td>
</tr>
<tr>
<td>Caropreso &amp; White, 1994</td>
<td>Gifted versus nongifted children, aged 4-6, favoring the gifted</td>
</tr>
</tbody>
</table>

Johnson (1991, 1996) has asserted the primacy of both linguistic ability and language-related knowledge in interpreting differences in metaphor development, with
findings supporting linguistic ability first and language-related knowledge in certain
cases. Jones and Stone (1989, as cited in Johnson, 1991) found lower performance on
tests of metaphor comprehension in learning disabled versus normally-functioning
adolescents, which they attributed to the former's limited capacity to know and use the
more subtle, connotative meanings of words.

Beyond studies focused specifically on linguistic connections, however, others
have found results related to a broader range of ability domains. Johnson's studies (1991,
1996; Johnson & Pascual-Leone, 1989) found high correlations between performance on
measures of metaphor interpretation and on a nonverbal measure of mental capacity.
Although in the same set of studies, IQ scores did not consistently show significant
correlations with metaphor performance, the researchers again emphasized the key role of
age-related differences: a calculated mental age capacity did consistently correlate with
performance on the metaphor interpretation measure. They explained the discrepancy
with the observation that IQ is a fixed variable, while development is dynamic, and that
children who were advanced in their development were more readily able to interpret
metaphors. Fung (1995) explored the influence of differences in cognitive functioning as
related to the role of context in children's metaphor comprehension at kindergarten and
grade three, finding a similarly inconsistent pattern related to IQ: this study found
evidence that cognitive differences related to metaphor comprehension among the third
graders but not among the kindergartners.

Castillo (1998) conducted a study that did not compare differences between gifted
and nongifted children but nevertheless examined key variables in metaphor
comprehension. She examined the effects of analogy instruction on metaphor
comprehension in young gifted children (ages 5.6 to 6.6), finding that instruction in using the form of an analogy to support understanding a metaphor led to positive effects on student performance when compared to that of a control group. Moreover, Castillo used two levels of scaffolding and instruction with the students, finding that those students exposed to additional mediation as they learned the analogy format outperformed those whose mediation was more limited.

Another study of the differences between performance in gifted and nongifted young students is related to this discussion by extension. Analogical reasoning is an important foundation of metaphoric competence (Castillo, 1998; Vosniadou, 1995), because as mentioned previously, an analogy that involves a figurative comparison is an expanded form of a proportional metaphor. The Castillo (1998) study mentioned above was a study of verbal analogical reasoning and its relationship to metaphor comprehension. Lohman (2001) and others (e.g., Chi, Hutchinson, & Robin; Perkins & Salomon, 1989) have discussed the transferability of some general cognitive processes across domains. If the processes involved in analogical reasoning are assumed to be transferable, then examination of geometric analogy studies has relevance to the discussion as well. Caropreso and White (1994) studied the differential performances of gifted and nongifted children between the ages of four and six on the Test of Analogical Reasoning in Children, which includes a series of geometric analogy problems. Children in this study were identified as gifted or nongifted based on their schools’ identification models, which included creativity and standardized intelligence measures as well as teacher recommendations. The results of this study demonstrated a significant difference in performance in analogical reasoning tasks between gifted and nongifted young
children, favoring the gifted group. Moreover, additional analyses indicated that among the gifted group, there were no significant differences based on gender or socio-economic status. From these results, the researchers concluded that the test used was a good identifier of giftedness in young children, and they suggested that analogical reasoning strategies could form an appropriate part of a strong curriculum for young gifted children.

In the present study, these findings around ability-related differences in performance on measures of figurative competence lent support to the focus on young, potentially gifted children as an appropriate population of focus for the instructional intervention. As a backdrop, then, to the particular population sampled for the study, a discussion of literature related to young gifted children follows.

*Young Gifted Children*

Much of the research about young gifted children has focused on their characteristics and how they may be identified from the larger population of students in the preschool and primary years. In many respects, young gifted children possess some of the same traits as older gifted children, notably in that their development is significantly more advanced than that of their age peers in given areas. The areas in which young gifted children have been observed to demonstrate advanced development include early and advanced language development, early facility with numbers and with spatial tasks such as jigsaw puzzles, a precocious sense of humor, the ability to manipulate language for social purposes, and early abstract thinking (Lewis & Louis, 1991). In addition, young gifted children demonstrate great asynchrony with regard to their different developmental areas, frequently manifesting itself as motor development that does not progress as quickly as cognitive development, leading to great frustration particularly with tasks
requiring manual dexterity such as writing (Silverman, 1997). Congruent with this focus on characteristics of young gifted children, considerable research has been conducted on effective ways of identifying such children for interventions in schools. Among the types of identification measures found to be effective have been standardized intelligence and achievement tests (Borland, 1986; Robinson, Abbott, Berninger, & Busse, 1996); portfolio approaches (Borland & Wright, 1994; Coleman, 1994; Shaklee, 1992); and parent recommendations (Louis & Lewis, 1992).

Somewhat less research has been conducted on the effects of various interventions with young gifted children in the classroom setting. A few interventions in laboratory or special program settings have been investigated to determine effects on student learning, with results showing significant gains for young gifted children when given the support of instruction and scaffolding, even in areas of development commonly thought to be only within the purview of older individuals (Castillo, 1998; Kanevsky, 1990; Robinson et al., 1996). Furthermore, some of this research has shown that although gains might be seen through instruction with nongifted children as well, the gains tend to be significantly greater for gifted children (Kanevsky, 1990). These studies have included attention to problem-solving strategies (Kanevsky, 1990); mathematical skills (Robinson et al., 1996); and metaphor comprehension (Castillo, 1998). Each of these areas, including metaphor comprehension as discussed above, depends upon student reasoning skills, which reflect a combination of generalized strategy as well as domain-specific knowledge (Perkins & Salomon, 1989); recent examination of theories of intelligence suggest that inductive reasoning abilities are at the core of general intellectual ability (Lohman, 2001), from which one might extrapolate that instruction related to these areas
of reasoning might have an even more powerful effect on gifted students than on their peers because of readiness. However, little information is available as to how well such interventions work in the classroom setting (Shore & Kanevsky, 1993).

Data on how well young gifted children are being served in their regular school settings is primarily anecdotal and contained within recommendations for more differentiated service for young gifted children in the classroom (Shaklee, 1992; Smutny, 1998). However, evidence from studies of classroom practice with older gifted children indicates that in regular classroom settings, little to no differentiation takes place for gifted students (Archambault et al., 1993; Westberg et al., 1993). Moreover, surveys of gifted children at elementary, middle, and secondary levels have chronicled their feelings that they are unchallenged by their instruction in the regular classroom (Gallagher, Harradine, & Coleman, 1997). No evidence exists to suggest that service to young gifted children is any more differentiated than that for their older counterparts; indeed, many suggest that the situation might be worse for young gifted children because of the unwillingness or inability of school programs to serve gifted children in the primary grades (Karnes & Johnson, 1987, 1991). Furthermore, although current recommendations in early childhood education are for "developmentally appropriate programs" that serve the individualized needs of all young children (Bredekamp & Copple, 1997), research into how differentiated and individually appropriate early childhood classrooms are tends to show an abundance of whole group instruction at the expense of differentiation for children at any level of exceptionality (Bryant, Clifford, & Peisner, 1991).

Because of the dearth of materials available to support differentiation for gifted children in the primary grades of school and because of their infrequent identification for
gifted programs (Barbour & Shaklee, 1998), a great need exists for the development of curricular materials that address advanced developmental needs. In the area of language arts, approaches to teaching literature that emphasize analysis and interpretation with targeted instructional moves on the part of the teacher have been found to support student growth in thinking (Baumann & Ivey, 1997; Beck & McKeown, 1999) and to lead to demonstrated growth in literary analysis and interpretation (VanTassel-Baska, Johnson, Hughes, & Boyce, 1996; VanTassel-Baska, Zuo, Avery, & Little, in press). For gifted students, recommendations for teaching language arts have included the use of complex and advanced reading materials as the basis for such approaches (Baskin & Harris, 1980; Nelson, 1990). Because of the inherent complexity and abstraction of the metaphoric structure and because of the potential for varied interpretations of figurative language, metaphors within quality works of literature may then be a strong basis for a curriculum strand for the gifted in the language arts. Thus, the following section will address existing research around instruction in metaphor as well as research on curriculum for the gifted, with an eye to the alignment of the two in the present study.

Strand IV: Figurative Competence and Instruction

Studies of Instructional Effects on Figurative Competence

As discussed previously, most of the studies of children’s metaphor comprehension and figurative competence have been laboratory-based studies that involved little or no instruction, focusing instead on children’s entering developmental levels and in some cases on familiarity with items. However, several studies, including some of the more recent work in metaphor development, have been conducted to examine the effects of instruction on various types of figurative competence. Feichtl (1988)
conducted a study in which elementary students participated in a course on proverbs and then were given two tests of figurative competence, with the finding that the treatment group outperformed the comparison group on both. In another study, Castillo (1998) investigated the effects of instruction in analogies on young children's comprehension of metaphors. Arguing that "analogies are the logical underpinnings of metaphors" (p. 28), Castillo used analogy instruction to help young gifted children understand proportional metaphors (analogies with one term left unstated, e.g., "jam is paint for bread") and found significant differences on a metaphor measure between the group exposed to analogy instruction and a control group. Each of these studies focused the instruction on a particular type of figurative language and then tested a broader range, concluding that the instruction had an effect on broader figurative competence.

Rudden (1994) and Redmond (1997) investigated the effects of metaphor instruction and figurative stimuli, respectively, on elementary school children's use of figurative devices in their own writing. In each case, the researchers found that the instruction seemed to have a positive effect on students' ability to use such devices and to increase the complexity of their use. Rudden (1994) asserted that the instruction had the effect of moving children from a more literal stage of expression into a more advanced level of metaphoric competence.

*Cognitive Mediation and the Zone of Proximal Development*

This notion of utilizing a structured instructional process to help children in moving from a less advanced to a more advanced level in a given area of development relates to Vygotsky's (1978, 1986) notion of the zone of proximal development. One of the implicit principles of education is the idea that the process of learning can be
enhanced by the guidance or mediation of others. Beyond this basic idea, however, some psychologists have theorized that the actual process of development can be enhanced and accelerated by the intervention of supportive teaching. Vygotsky's theory of the zone of proximal development suggests that children actually have two different levels of development in a given domain at a given time: one represents their level of development as they work independently, and the other represents their level of development when given assistance or mediation. The region between these two levels is termed the zone of proximal development, and Vygotsky argued that a child's developmental level could be raised from the independent level closer to the mediated level through concentrated instruction and assistance. Furthermore, instructional activities within this zone of proximal development are most likely to provide the "optimal match" that is the hallmark of effective teaching – posing problems that are not so easy that they cause boredom nor so difficult that they cause frustration (Benbow, 1998; Csikszentmihalyi, 1990), and using mediation as a support structure for maintaining progress in the zone (Feuerstein, Rand, Hoffman, & Miller, 1980).

Karpov and Haywood (1998) discussed Vygotsky's views on the zone of proximal development and on cognitive and metacognitive mediation and how they have been translated into instructional practice. According to their discussion, guided discovery programs that encourage students to work in groups to construct their understanding of given topics and concepts (e.g., Brown & Campione, 1994) reflect Vygostky's recommendations for metacognitive mediation, in which students learn processes for monitoring and evaluating their own learning processes, but less so his notions of cognitive mediation, which require direct teaching of what Vygotsky referred
to as "scientific concepts." Vygotsky’s perspective was that students needed to be
directly taught the scientific concepts undergirding the various disciplines, although not
in a way that caused them only to memorize vocabulary by rote. Rather, cognitive
mediation requires that teachers give students direct instruction around given concepts,
with scaffolding to encourage not only the memory of the concepts but their appropriate
application in various contexts. As Bruer (1993) noted, students who are guided by
cognitive mediation “are taught methods of scientific analysis aimed at identifying and
modeling the essential characteristics of objects and events. Having been internalized,
these methods become cognitive tools that mediate students’ independent problem
solving” (p. 33). Karpov and Haywood (1998) recommend a combination of this focused
cognitive mediation with follow-up group discovery supporting metacognitive mediation.

*Individual Differences and the Zone of Proximal Development*

Vygotsky (1987) noted, in cautioning readers about overgeneralizing the zone of
proximal development, that

> We said that in collaboration the child can always do more than he can
> independently. We must add the stipulation that he cannot do infinitely more.
> What collaboration contributes to the child’s performance is restricted to limits
> which are determined by the state of his development and his intellectual potential
> (p. 209; quoted in Daniels, 1996, p. 15).

Looking at this from a slightly different perspective, researchers in the field of gifted
education have worked under the presumption that gifted children, often already at a
higher level of development than their age peers, also have a wider zone of proximal
development, or greater learning potential, through which they may be guided with the
assistance of a teacher. Studies examining the performance of gifted and nongifted young children with guidance from an instructor have shown that gifted children make greater gains than their nongifted peers, and that they use guidance from an instructor more efficiently in their learning process (Kanevsky, 1990). Similarly, the idea of individualized attention planned to meet specific developmental needs is a foundation for mentorship programs, which have long been used with advanced students to challenge and motivate their learning in specific domains of interest and ability (Clasen & Clasen, 1997). Indeed, Clasen and Hanson (1987) suggested that successful mentorships are often the result of careful attention to developmental needs of the student, with attention to mediation and instruction above the boredom level and below frustration level. Such a focus also lends itself to giving students opportunities for “flow,” or the experience of being fully engaged to the exclusion of all distraction with a task that is interesting and developmentally appropriate (Csikszentmihalyi, 1990).

Metaphor Development and the Zone of Proximal Development

Vygotsky’s notions of the zone of proximal development and of targeted cognitive mediation may be applied specifically to the inclusion of metaphor as an element of a language arts curriculum. With regard to mediation, the specific elements of metaphor structure, along with the underlying process of analogical reasoning as a component of metaphor comprehension and production, represent an organized scaffold that may be taught to students as a support for understanding metaphors they encounter and recognizing and refining them within their own writing efforts. Furthermore, if exposure to examples of metaphors is linked with study of the theoretical foundation of metaphor, children have the opportunity to use metaphors to find out about new concepts.
in a discovery-oriented process (Brown & Campione, 1994) while using the definitional structure of metaphor as a metacognitive mediator (Karpov & Haywood, 1998) to support their understanding of the differences between conceptual categories. Consequently, the specific processes related to metaphor comprehension may be taught to students and then taken by them as a tool for approaching and evaluating their more independent work with metaphor in their reading and writing. Moreover, direct teaching of a way of analyzing new conceptual comparisons can support students in their further learning in multiple areas of study.

Instruction in metaphor also has a broader significance, beyond just how it can be taught as an individual area of study; it also has an important place within the larger scope and sequence of language arts instruction. Steinbergh (1999) quoted teachers who asserted that students who were not well versed in metaphor interpretation by the time they reached high school level English courses would be lost in their attempts to read and analyze the works of literature included in such courses. Certainly by high school but also by earlier in their language arts scholastic career, students will be exposed to metaphor frequently in the literature they read, and their understanding of that literature can become much broader and richer if the understanding of metaphor is more solidly grounded. Moreover, teaching metaphor by its structure early in a language arts program can provide a scaffold for the creative writing opportunities that are so often provided but not always well structured or guided in schools.

The zone of proximal development and the notions of cognitive and metacognitive mediation are important to consider in this discussion of metaphor development because of the variability in children’s development of figurative
competence, and indeed, of literacy broadly defined (Tierney, 1991; Vosniadou, 1987). Because children progress through the developmental processes at different rates and thus may benefit to different degrees from targeted instruction in metaphor, such instruction has the potential to serve as appropriate differentiation for those students who are advanced in their metaphor development and ready to move to another stage of figurative competence. Furthermore, since NAEYC's guidelines for developmentally appropriate practice support individualized attention based on developmental level (Bredekamp & Copple, 1997), such differentiation is also in line with recommendations of the early childhood community.

Although metaphor development is a process common to all growing children, and although metaphors will be encountered by each child in various communicative circumstances, metaphor instruction because of its differential levels of complexity and because of the abstract reasoning that underlies it may provide a specific opportunity for appropriate instruction for young children with higher levels of cognitive ability than their age peers. Thus, the present study utilized an intervention designed around research-based principles of curriculum for the gifted and employed a sampling strategy intended to include children potentially gifted in the verbal areas. The following section will outline research related to curriculum for the gifted as a foundation for the design of the intervention utilized in the study.

Curriculum for Gifted Students and the Integrated Curriculum Model

In a review of effective program practices in gifted education, Shore and Delcourt (1996) found that the use of high-level curricular materials was among five practices uniquely appropriate to gifted education. Kulik and Kulik (1992) also found that
differentiation of curriculum was the key variable in terms of the effectiveness of ability grouping for both gifted and non-gifted students. However, in a discussion of existing research on curriculum for the gifted, Johnsen (2000) noted that although “curriculum and instruction are the heart of education for gifted and talented students” (p. 25), there have been only a limited number of studies exploring effects of curricular interventions with this population. Similarly, reviews of general curriculum materials have shown that few textbook programs, particularly in the language arts, are appropriate for gifted students, based on criteria related to characteristics of the population (Aldrich, 1996).

Several types of curriculum interventions have proved to be effective with gifted students across a range of ages and curricular areas. Content acceleration, particularly in mathematics and science, has a long history of effectiveness in terms of continued high achievement levels among students who have participated in targeted acceleration programs (Kulik & Kulik, 1992; Lynch, 1992; Swiatek & Benbow, 1991). Problem-based learning, which engages students in exploring ill-structured, real-world problems and learning content as a consequence of their problem exploration, has also been found to be effective for gifted students in terms of their problem finding and solution finding skills (Gallagher, Stepieen, & Rosenthal, 1992), and a study of student achievement in experimental design within a problem-based science unit also found significant improvement for gifted students as compared to a comparison group (VanTassel-Baska, Bass, Ries, Poland, & Avery, 1998). Another study investigating the cognitive level of teacher questioning and student responses found a strong positive relationship between the two – higher-level questioning elicited higher-level responses from students (Friedman & Lee, 1996). Each of these studies demonstrates the importance of planned
curriculum and teacher involvement – in the case of content acceleration, to support students in developing understanding at rapid rates beyond their current achievement level; and in the case of problem-based learning, to facilitate the problem investigation process and ensure that materials and resources are in place to support content learning. The study of questioning demonstrates the importance of preparing questions that are at a high cognitive level and supporting students in developing the habits of mind that enable them to respond to such questions.

VanTassel-Baska (1986, 1994) discussed three dimensions of curriculum models employed to serve the academic needs of gifted students: content models, such as those used in acceleration programs, which focus on moving students rapidly through the levels of content in particular domains; process/product models, which teach students higher-level thinking and involve them in developing advanced products, often through independent study in areas of interest; and concept models, which use interdisciplinary and abstract concepts as the organizers for curriculum. Each of these categories relates specifically to a critical element of academic talent. Gifted students tend to be precocious in their learning; thus, advanced content may serve their needs. Often, particularly in areas of interest, gifted students show intensity in their ability to focus on a high-level task for an extended period; thus, process/product models may encourage them to continue to develop this habit of intensity. Finally, because of the complexity gifted students are capable of showing in their thinking, concept models that encourage them to explore abstract ideas at great depth provide opportunities for growth.

The Integrated Curriculum Model (ICM) (VanTassel-Baska, 1986, 1995) is a foundation for curriculum that integrates features of all three of the model dimensions
described above – content, process/product, and concept. The model is also based on an array of curriculum reform elements relating to trends in education broadly defined and to the national standards projects (e.g., Rutherford & Ahlgren, 1989; National Council of Teachers of English/International Reading Association, 1996). Among these elements are learner outcomes of significance, inquiry-based learning, conceptually-oriented curriculum, and engaging students in the processes of constructing meaning and practicing metacognition (VanTassel-Baska, 1995).

Several series of curriculum units have been developed based on this model, including units serving gifted students at various grade levels in science, language arts, and social studies. Research on the units in science and language arts has demonstrated significant learning gains for students as compared to a comparison group not receiving instruction in the units (VanTassel-Baska, Bass, Ries, Poland, & Avery, 1998; VanTassel-Baska, Johnson, Hughes, & Boyce, 1996; VanTassel-Baska, Zuo, Avery, & Little, in press). Moreover, teachers and students who participated in the units have described benefits relating to motivation, interest, and developing strong reasoning skills and habits of mind relating to the disciplines under study (VanTassel-Baska, Avery, Little, & Hughes, 2000). In the language arts area specifically, the effectiveness of the units with regard to student growth in literary analysis and interpretation (VanTassel-Baska, Johnson, Hughes, et al, 1996; VanTassel-Baska, Zuo, et al., in press), demonstrates potential for further work in the processes of developing conceptual understanding through text.

Comprehension of metaphor requires that a reader or listener analyze the content of the figurative statement and its context, realize that it is meant to be interpreted
figuratively, and then reach an understanding of the comparison involved (Vosniadou, 1987). This requires an advanced level of analysis and interpretation, along with an ability to reason analogically (Castillo, 1998), and a level of conceptual language development that acknowledges categories of concepts and intentional violation of them (Levorato, 1993). In order to meet these requirements, a curriculum unit designed around metaphor must present students with literature that includes figurative comparisons in context, introduce them to the process of analogical reasoning, and encourage them in their developing understanding of the concept of language and how language may be changed to suit the needs of the speaker or writer. The ICM was thus used to develop the intervention unit for the proposed study, with advanced content incorporated through advanced literature, higher level process through analogical reasoning, and concept development through the integrated study of the themes of language and change.

Conclusion

Figurative language is a constant part of our everyday life and speech; metaphors enrich language aesthetically and provide effective means of learning about new concepts based on relationships to the familiar. Figurative competence is a developmental process, related to overall linguistic development and to language experience. The role of instruction in the process of developing figurative competence is still unclear because of limited study; however, the research does indicate that as children progress through the primary grades of school, they pass through a literal phase into a level of development at which they begin to suspend the literal and come to understand and appreciate the figurative (Levorato, 1993). Consequently, the primary years represent an opportunity to encourage the developmental process by providing exposure through literature-based
experiences to figurative language, and by creating scaffolds that will assist students in understanding the metaphors they encounter and in devising their own metaphorical expressions. Such a literature-centered and language experience-based approach to teaching metaphor not only reflects theories of cognitive psychology about how children learn in general (Vygotsky, 1978, 1986), but also relates to the prevalent literature-based and skill-balanced approach to teaching early elementary language arts that has emerged under the emergent literacy paradigm (Calfee & Norman, 1998).

For children who are advanced in their development above their age peers, the conceptual complexity of metaphor and the analogical reasoning processes involved in metaphor comprehension represent advanced language content that can challenge them and provide differentiated learning opportunities in the classroom, even though these students may not be officially identified as gifted. The issue of providing appropriately differentiated learning opportunities for advanced students in the regular classroom is an important one that is raising concerns across the grade levels (Archambault et al. 1993; Barbour & Shaklee, 1998; United States Department of Education, 1993); thus, exploring options for curricular interventions that are effective with the population and usable in the regular classroom is an important direction for research and practice.

This study examined to what extent an instructional unit designed under an established curriculum model for gifted students, with content based on psycholinguistic principles of the nature of metaphor, affected young children's metaphor comprehension and figurative competence. The next chapter will outline the methodology utilized in implementing the study.
CHAPTER III

Methodology

Purpose and Research Questions

The purposes of this study were to explore the developmental readiness of young, verbally talented students to engage in tasks related to the comprehension and production of metaphor, and to determine the effects of an intervention, organized around the principles of metaphor, on second grade students' performance on measures of figurative competence. A further purpose of the study was to determine to what degree said performance related to student performance on standardized measures of verbal ability.

The following research questions formed the foundation of the study:

1. Do student scores on measures of figurative competence relate significantly to scores on standardized ability and achievement tests in the verbal areas and in general cognitive ability?

2. To what extent are abilities related to figurative competence demonstrated in student products completed during an instructional intervention focused on metaphor?

3. Does instruction in the definitional structure of metaphor and use of semantic mapping make a significant change in performance on measures of figurative competence assessing metaphor comprehension and literary analysis and interpretation in second grade, verbally advanced students?

4. To what extent does student performance on written measures of figurative competence relate to teacher evaluation of overall student performance on written and oral tasks completed during the intervention?
The hypotheses for the research questions were as follows:

1. There will be a significant positive correlation \((p < .05)\) between student scores on measures of figurative competence and their scores on standardized ability tests in the verbal areas and in general cognitive ability.

2. Evidence of figurative competence meeting and exceeding age-based expectations will be demonstrated in student products completed during the intervention.

3. Significant differences \((p < .05)\) will be evident from pretest to posttest performance on tests of metaphor comprehension and literary analysis and interpretation among students in the treatment condition as compared to students in the comparison condition, with students in the treatment condition demonstrating greater gains than those in the comparison condition.

4. Student performance on written measures of figurative competence will show a significant positive correlation \((p < .05)\) to teacher evaluation of overall student performance in written and oral tasks engaged in during the intervention.

**Conceptual Framework**

The conceptual framework for the study was based primarily on Levorato's Model of Figurative Competence (Levorato, 1993; Levorato & Cacciari, 1995), as explained in Chapter II and demonstrated graphically in Figure 1. According to this model, figurative competence is linked to the development of a broader spectrum of linguistic skills, important in its contribution to an overall semantic competence and language comprehension. The model identifies a series of six developmental levels through which an individual passes in the process of moving from a "nominal realist
phase," in which objects and their names are seen as the same thing, to "metalinguistic competence," or the stage at which language facility and comprehension are overlaid with an advanced level of reflection upon language itself (Levorato, 1993, p. 119). This achievement is characterized by a set of six abilities relating to developing understanding of varied word meanings and usages; analyzing relationships between literal and figurative uses of words; and processing and producing large amounts of language as figurative units. These six abilities are specifically outlined in Chapter II and in Figure 1. Figure 1 also demonstrates how the specific abilities may be related to the levels of the model, based on descriptions of behaviors characterizing the different levels as described by Levorato (1993; Levorato & Cacciari, 1995).

This study, with this model as a framework, also utilized Vygotsky’s (1978, 1986) theory of the zone of proximal development, which postulates that in a given domain, an individual is capable of functioning at a certain developmental level on his own, but can function at a more advanced level with assistance through scaffolding around the content and process under study. The range between performance level alone versus performance level with mediation is the zone of proximal development. Scaffolding, in Vygotsky’s framework and in the work of more contemporary Russian educators, is designed to be a manifestation of both cognitive and metacognitive mediation, giving students direct guidance and instruction around specific concepts and related processes as well as tools for independent, self-monitored application of said processes (Karpov & Haywood, 1998). Within the context of this study, the global scaffolding structure was the intervention unit, with direct teaching of the elements of metaphor and opportunities to
apply analysis of these elements across a range of literary inputs as the specific
components of the unit forming the building blocks of the scaffolding structure.

The components of the intervention that represent cognitive mediation focused on
the knowledge, abilities, and behaviors that characterize Levels 1 through 3 of the Model
of Figurative Competence, which Levorato (1993) has indicated as the levels of
elementary-age children. Specifically, the intervention focused on the abilities related to
understanding the dominant, peripheral, and polysemous meanings of words; suspension
of a literal strategy; and the ability to understand figurative uses of a word and the
relationship between literal and figurative meanings. This focus was achieved through the
deliberate design of the intervention to include the specific figurative ability emphases as
it explores the idea of metaphor and several types of figurative language. The study also
involved opportunities for exploration of abilities and behaviors characteristic of the
more advanced ability levels by providing metaphor production opportunities, with the
assumption that gifted children may demonstrate abilities well beyond the developmental
level of their age peers (Jackson & Klein, 1997).

Research Design

This study involved a mixed design, utilizing both quantitative and qualitative
approaches. The primary emphasis was on the quantitative methods, utilizing a quasi-
experimental, non-equivalent control group design (Gall, Borg, and Gall, 1996), where
the control group was actually a comparison group receiving a different treatment rather
than a group receiving no treatment. The third research question required experimental
data, and the quasi-experimental design permitted data-gathering and inferences from
existing students and classes, which placed the study in the real-world context of schools.
Pre- and post-assessment data were utilized to allow measurement of change and to control for group differences between the treatment and comparison groups. Qualitative methods were used for data collection and analysis around Research Question 2, which allowed elaboration about students' demonstrations of figurative competence beyond the quantitative data related to the other questions.

Research Questions 1 and 4 required a correlational research design, in order to assess relationships between different measures of ability and performance both within and outside of the intervention context (Gall, Borg, & Gall, 1996). Again, the use of pre- and post-assessments allowed correlations to be conducted regarding change in performance rather than solely around a single measure.

Table 2 outlines the four research questions with the relevant data sources and analysis techniques used for each. Methods for data analysis are also discussed in additional detail later in the chapter.
<table>
<thead>
<tr>
<th>Question</th>
<th>Data Source(s)</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Do student scores on measures of figurative competence relate</td>
<td>• Posttest and gain scores on MCT and LAIT</td>
<td>• Multiple regression</td>
</tr>
<tr>
<td>significantly to scores on standardized ability and achievement tests in the verbal areas and in general cognitive ability?</td>
<td>• TCS scores</td>
<td></td>
</tr>
<tr>
<td>2. To what extent are abilities related to figurative competence</td>
<td>• Student products</td>
<td>• Content analysis for</td>
</tr>
<tr>
<td>demonstrated in student products completed during an instructional intervention focused on metaphor?</td>
<td></td>
<td>patterns, based on case study methodology</td>
</tr>
<tr>
<td>3. Does instruction in the definitional structure of metaphor and use of semantic mapping make a significant change in performance on measures of figurative competence assessing metaphor comprehension and literary analysis and interpretation in second grade, verbally advanced students?</td>
<td>• Pre-post test scores on MCT and LAIT</td>
<td>• ANCOVA between treatment and comparison groups</td>
</tr>
<tr>
<td>4. To what extent does student performance on written measures of figurative competence relate to teacher evaluation of overall student performance on written and oral tasks completed during the intervention?</td>
<td>• Pre-post test scores on MCT and LAIT • TASP scores</td>
<td>• Descriptive comparisons • Correlations (Pearson) between TASP and MCT/LAIT scores</td>
</tr>
</tbody>
</table>
Sample

The study sample included students designated to be at advanced reading levels in six classes of second grade students across six schools in one large, urban-to-rural school district, and one class of advanced second grade readers in a small suburban school district. Five of the six classes in the larger district served as treatment group classes, while the remaining class in the large district and the single class in the smaller district served as comparison classes. The district from which the majority of the sample was drawn has an overall enrollment of nearly 135,000 students in 189 schools, making it the 19th largest school system in the United States, with a diverse student population in terms of ethnic and socioeconomic background (Montgomery County Public Schools, 2000). The additional class in the comparison group came from a district with 12 schools. This second district includes suburban and rural areas and a population diverse in terms of ethnic and socioeconomic background, although it is somewhat less diverse than the study district, primarily because of size and location. The seven schools and the context of their participation in the study are summarized in Table 3.
Table 3

**Schools Included in Sample: Treatment Group**

<table>
<thead>
<tr>
<th>School</th>
<th>No. of Students</th>
<th>Teaching Assignment/Study Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>School 1</td>
<td>12</td>
<td>Reading Initiative teacher/ Daily language arts pullout</td>
</tr>
<tr>
<td>(Primary</td>
<td></td>
<td>Gifted Magnet)</td>
</tr>
<tr>
<td>School 2</td>
<td>9</td>
<td>Classroom teacher/Top reading group language arts</td>
</tr>
<tr>
<td>School 3</td>
<td>23</td>
<td>“William and Mary” teacher assigned to W&amp;M language arts K-5/ Daily language arts pullout</td>
</tr>
<tr>
<td>School 4</td>
<td>17</td>
<td>Classroom teacher/ Top reading group language arts</td>
</tr>
<tr>
<td>School 5</td>
<td>9</td>
<td>Gifted Resource Teacher/ Daily language arts pullout</td>
</tr>
</tbody>
</table>

**Schools Included in Sample: Comparison Group**

<table>
<thead>
<tr>
<th>School</th>
<th>No. of Students</th>
<th>Teaching Assignment/Study Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>School 6</td>
<td>12</td>
<td>Reading Initiative teacher/ Daily language arts pullout</td>
</tr>
<tr>
<td>School 7</td>
<td>9</td>
<td>Gifted Resource Teacher/ Twice weekly gifted pullout</td>
</tr>
<tr>
<td>(from different district)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All of the teachers in the treatment group were trained in using the intervention. Within the school district, all second grade teachers were offered the opportunity to participate in the voluntary training and to try out the unit in their own classrooms as part
of a language arts initiative to use units incorporating the ICM across all elementary
schools in the county.

Following the training workshops and after approval of the study by the
researcher's dissertation committee, all teachers who attended the workshop were given
the option of participating in the study as either treatment or comparison group teachers.
The parameters of study involvement were outlined in detail, in written form, for teachers
to consider. Although approximately 100 teachers were involved with the training
initiative, only six of these agreed to participate in the study. Among these six, one
teacher agreed to participate only in a comparison class capacity, with the understanding
that she would not employ any of the unit lessons until after the data collection period.
The other five volunteers agreed to serve as treatment classes. One additional comparison
teacher was found on the second grade staff at one of the treatment class schools.

At the conclusion of the study, teachers were asked to complete a brief email
questionnaire regarding their experiences with the intervention and also to provide some
background data on their education and experience. These descriptive details on the study
teachers are reported in Chapter IV.

Within the classes, students who participated in the study were selected based on
their scores on a district test of reading fluency and comprehension. This test, a
performance-based reading assessment administered to all second grade students at the
beginning of the school year, assesses student literacy level for use as a basis for
grouping and instructional differentiation in reading. At grade 2, the test assesses whether
students are fluent in reading text measured at a second grade readability and whether
they are able to respond correctly to comprehension questions based on the reading.
Students who exceeded the district’s grade 2 expectations were judged to be reading above grade level and were eligible for participation in the study. Within the sample, student performance on the test ranged from only slightly above grade level (one-half to one year) to several years above grade level, but no specific reporting of student performance on this test was requested.

The sampling procedures for the study resulted in a treatment group of 70 students and a comparison group of 21. Several attempts were made to increase the size of the comparison group but with no results. Students in the treatment group ranged in age from 7 years 1 month to 8 years 3 months, with a mean age of 7 years 10 months. The comparison group represented a slightly older range, from 7 years 4 months to 8 years 6 months, but the mean age for the comparison group was also 7 years 10 months. Ages are reported based on time of pretest administration. The treatment group included an even distribution of males and females, 35 of each gender; the comparison group included 12 females and 9 males.

Near the conclusion of the study, screening for the district gifted and talented program was conducted across the second grade classes. Scores on the screening instruments were provided to the researcher for students in the study sample; relevant sample demographics regarding these scores are reported in Chapter IV. Additional demographic data on the students, including ethnic background and free/reduced meal status, were requested, but the school district did not permit their release on an individual basis; thus, only school-wide demographic data were available and are included in the school descriptions in Appendix A.


Instrumentation

The study employed four specific instruments to gather data related to students’ figurative competence, as well as analysis of student portfolios. The instruments included the following: (a) pre- and post-tests of metaphor comprehension, (b) pre- and post-tests of literary analysis and interpretation, (c) a teacher assessment of student performance, and (d) a classroom observation scale. Each of these instruments was designed or modified from existing instruments by the researcher, and each was piloted prior to use in the study. The instruments, with details on piloting results, are discussed in the sections that follow.

Metaphor Comprehension Test (MCT)

The test used to assess metaphor comprehension in the study was a multiple-choice instrument, designed by the researcher, that measured student ability to recognize and comprehend metaphors in context (see Appendix B). The assessment was designed to reflect several of the specific factors potentially influencing performance on metaphor comprehension tasks with young children, as outlined by Vosniadou (1987). First, since the assessment is composed of multiple-choice items, it allows children to demonstrate levels of comprehension without having to produce linguistically complex explanations in writing themselves. Second, the assessment asks questions about metaphors found within given poetry selections; thus, the children are reading the metaphors within their intended context, and the context can thus support recognition of figurative instead of literal intent.
Test specifications.

Two forms of the test were developed, each containing four different poetry selections. Each poem was followed by two to four multiple choice questions regarding the metaphors contained within the text, for a total of 13 items per form. The poems were selected based on three primary criteria. First, they needed to contain at least one metaphor, with both topic and vehicle stated somewhere within the poem, although the specific form of the metaphor varied from poem to poem. Second, the poems needed to be short, so that the overall assessment period, including reading and response to questions, would be brief. Third, the vocabulary contained within the poems needed to be reasonably simple, to the degree that most second graders reading at least on grade level would be able to read and understand the words without assistance.

The test items were designed around the definitional structure of a metaphor: that is, that a metaphor is composed of a topic, a vehicle, and a ground, and that its intent is to communicate emphasis on specific characteristics of the topic. Thus, test items assess children's ability to recognize these components of a metaphor in context and to identify the salient characteristics of both topic and vehicle, as well as to demonstrate comprehension of the overall meaning of the metaphor. Table 4 illustrates the specifications of the test items as they relate to expected student performance around the content (Gronlund, 1998).
### Table 4

*Item Specifications for Metaphor Comprehension Test*

<table>
<thead>
<tr>
<th>Task Demand</th>
<th>Number of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifies topic of metaphor</td>
<td>2</td>
</tr>
<tr>
<td>Identifies topic of a given vehicle</td>
<td>2</td>
</tr>
<tr>
<td>Identifies key characteristics of topic</td>
<td>3</td>
</tr>
<tr>
<td>Identifies vehicle of given topic</td>
<td>2</td>
</tr>
<tr>
<td>Identifies key characteristics of vehicle</td>
<td>3</td>
</tr>
<tr>
<td>Identifies main idea of selection</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note.* Item type distribution is the same for both forms of the instrument.

Scoring for the test was determined based on item difficulty according to the theoretical basis for how children develop metaphor comprehension and based on the availability of literal versus figurative responses among response choices. Two categories of items were designed as easier than the other four, and thus had a lower score possible: identifying the topic of the metaphor and identifying the vehicle of a given topic were simpler items, requiring only that students recognize the main thing that the poem was about and the main thing that item was compared to. Thus, each of these items had a total possible score of 1. The other four items required more sophisticated interpretation and thus had a total possible score of 2. This led to a total possible score of 22.

For some items, response choices included a correct figurative interpretation of the metaphor in question but also a literal interpretation that would have been correct had
literal interpretation been the intent. In cases in which this occurred, the literal response was given a score of one-quarter of the correct figurative response, so that any existing patterns regarding the literal to figurative stages of development could be revealed. In other cases, all other responses were incorrect and received a score of zero.

The two forms of the instrument were reviewed by committee members, revised, and then piloted with a relevant group of children, including review for aspects of content validity and readability, as well as testing for reliability (Janda, 1998). As a preliminary step, nine second grade students tried out both forms of the test and were told to ask for clarification on any items they found confusing. This tryout phase revealed no items needing rewording or clarification; discussion of test items demonstrated that any difficulty students experienced with the items was a result of being unsure as to the correct answer to the question rather than uncertainty as to what the question was asking. Thus, the scores from this tryout phase were also utilized in the determination of equivalent-forms reliability. Another set of students participated in the pilot to determine test-retest reliability, taking the same version of the test twice with an interval of two weeks between administrations.

**Equivalent forms reliability.**

The pilot group to determine equivalent-forms reliability consisted of 20 second grade students, ranging in age from 7 years 3 months to 8 years 6 months with a mean age of 7 years 9 months. The group included 11 boys and 9 girls. The group included the initial tryout group, nine students who took the tests in groups of two or three, and then a second group of 11 students who took the tests in a classroom setting. To account for potential fatigue and practice effects, the tests were administered in varying order; 11
students took form A followed by form B, while the remaining 9 took form B followed by form A. The correlational analysis between the two forms yielded a statistically significant correlation coefficient of .502 ($p < .05$). Table 5 gives the means and standard deviations of both forms for the pilot group.

**Table 5**

*Equivalent Forms Pilot Results (MCT)*

<table>
<thead>
<tr>
<th>Form</th>
<th>$M$</th>
<th>$SD$</th>
<th>$n$</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form A</td>
<td>12.23</td>
<td>4.24</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Form B</td>
<td>12.05</td>
<td>2.89</td>
<td>20</td>
<td>.502*</td>
</tr>
</tbody>
</table>

$*p < .05$

In order to account for the relatively low correlation between the forms, student scores were examined individually. No patterns of practice effect, gender difference, or grouping (between the small groups versus the classroom group) were evident in the three scores that varied widely from one form to the other, nor was there a pattern as to which test yielded higher scores among these students with such variance. One student whose behavior during test administration had demonstrated inattentiveness and a rush to complete the test had received scores that differed by 7.25 points; when his scores were removed and the data analyzed again, the correlation coefficient increased to a value of .582 ($p < .01$). In addition, two other students seemed to be outliers based on the difference between their form A scores and form B scores, with one student's scores varying by 8 points and the other by 6.25 points. All other student scores varied by 5 points or less from one form to the other. When the two students' scores were removed, in addition to the removal of the child who demonstrated inattention, the reliability
The coefficient increased to .672 ($p < .01$). The set of means and standard deviations for this analysis, which actually yielded more diverse means but closer standard deviations, may be found in Table 6.

Table 6

*Equivalent Forms Pilot Results (MCT), Outliers Removed*

<table>
<thead>
<tr>
<th>Form</th>
<th>$M$</th>
<th>$SD$</th>
<th>$n$</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form A</td>
<td>12.94</td>
<td>3.24</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Form B</td>
<td>12.31</td>
<td>3.06</td>
<td>17</td>
<td>.672**</td>
</tr>
</tbody>
</table>

*$$p < .01$*

In addition to the review of individual student total scores, the data were also analyzed to determine if any item pairs showed patterns of response that could account for the low correlation. Eleven of the thirteen item pairs showed similar response rates between the A item and its comparable B item; of the two remaining items, one indicated that the A item was more difficult, while the second indicated that the B item was more difficult, suggesting some measure of balance between the forms.

Because of the small size of the pilot group and the limited number of items on the test, individual students whose scores seemed to be outliers and individual items with varied responses between the A test and the B test had a large effect on the statistical results. However, the equivalent-forms testing did indicate a statistically significant correlation between the tests, and that correlation became stronger when students who were apparently outliers were removed from the analysis.
Test-retest reliability.

The test-retest reliability method was also utilized to assess the stability of the instrument (Janda, 1998). The two forms of the instrument were assessed separately for this purpose, with two different groups of students each taking one form of the test twice with a three-week interval between administrations. Each group included 10 students, all from the same school; 19 of the total 20 were second graders, with one highly able first grader included in the form A group.

The form A group included three boys and seven girls, ranging in age from 6 years 9 months to 8 years 2 months with a mean age of 7 years 10 months. The Pearson $r$ correlation coefficient showing the relationship between the two administrations of Form A was .862 ($p < .01$). The details of this analysis are shown in Table 7.

Table 7

<table>
<thead>
<tr>
<th>Administration</th>
<th>$M$</th>
<th>$SD$</th>
<th>$n$</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form A – first administration</td>
<td>16.03</td>
<td>3.12</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Form A – second administration</td>
<td>16.88</td>
<td>3.78</td>
<td>10</td>
<td>.862**</td>
</tr>
</tbody>
</table>

**$p < .01$**

The group taking form B of the instrument for test-retest reliability included 7 boys and 3 girls, ranging in age from 7 years 4 months to 8 years 4 months with a mean age of 7 years 9 months. The analysis of this group yielded a much lower correlation coefficient, with a value of .462, which was not statistically significant ($p > .05$). The results for this group are shown in Table 8.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Several reasons for the lack of a statistically significant correlation may be postulated here beyond the notion that the test is not stable. Of the ten students in this group, eight dropped in their scores from the first to the second administration, with only one student gaining in score and one student remaining the same. Thus, for this group of students, a fatigue or boredom with the test may have occurred. In the Form A group, on the other hand, seven of the ten improved their scores, and the differences between the two administrations was much smaller, suggesting possibly a slight practice effect. Moreover, as with the equivalent-forms reliability, the small size of the pilot group limits the reliability data because of the powerful influence of one participant’s score. However, also of some concern was the lack of noticeable variance in the scores; the limited sample size may have affected this as well, but the clustering of scores may suggest an instability in the instrument.

In addition to examining the pilot results for reliability, an effort was also made to examine the data for ceiling effect. Because the instrument is so short and because of the intended use with highly able populations, ceiling effect was an important possibility to examine. However, again because of the population taking the test and the manner of its development, the likelihood of a normal distribution of performance was slim, limiting

<table>
<thead>
<tr>
<th>Administration</th>
<th>M</th>
<th>SD</th>
<th>n</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form B – first administration</td>
<td>12.63</td>
<td>2.60</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Form B – second administration</td>
<td>11.07</td>
<td>2.12</td>
<td>10</td>
<td>.462</td>
</tr>
</tbody>
</table>
the degree to which statistical measures could be used to assess ceiling effect. Therefore, the scores were instead examined to determine how many students scored above a researcher-designated "cut-off" of 85% correct (L. Zuo, personal communication, February 2001). In the case of Form A, only one student exceeded this cut-off, with a score of 19.75 or 89.8% correct; on Form B, the highest score was 18, or 81.8% correct. Therefore, although the brevity of the test and the small pilot sample did not eliminate concerns about ceiling effect, the pilot data did not support the notion that ceiling effect was a significant problem with either form.

*Test of Literary Analysis and Interpretation (LAIT)*

A second set of pre- and post-assessments was also included in the study in order to get a different assessment of student performance with regard to metaphor comprehension. This instrument (see Appendix C) is a version of a test for literary analysis and interpretation, modeled on the National Assessment of Educational Progress instrument in reading (National Assessment Governing Board, 1992), that was piloted with groups of gifted students and used in earlier curriculum effectiveness work with language arts units developed under the Integrated Curriculum Model (VanTassel-Baska, Johnson, Hughes, et al., 1996; VanTassel-Baska, Zuo, et al., in press). In the context of this study, the test and rubric were modified slightly to account for the age level of the students and also to emphasize the related aspect of metaphor comprehension. This second assessment is included within the context of the intervention, and discussion of the pretest items and responses is incorporated into the lessons of the intervention.

For each version of this test, students are presented with a poem and asked to respond in writing to a set of four questions. The first question asks students to identify
the main idea of the poem; the second question asks them to explain a direct quote from
the poem; the third question asks students to identify the topic of the poem’s metaphor
and explain how its connection to the vehicle changed their thinking about the topic; and
the fourth question asks students to create a new title for the poem and to explain their
reasoning. The first and fourth question are intended to assess overall comprehension and
interpretation of the text, while the second and third focus more directly on the metaphor
aspect of the pieces. As in the case of the multiple choice test for metaphor
comprehension discussed above, the poems selected for the interpretation test were short,
had relatively simple vocabulary, and included at least one clear metaphor.

The rubric for this test was also utilized in prior studies, with interrater reliability
reported as .81 (VanTassel-Baska, Johnson, Hughes, et al., 1996) and more recently as
.90 (VanTassel-Baska, Zuo, et al., in press). The rubric was modified in this study to
reflect the emphasis on metaphor comprehension (see Appendix D). The rubric gives a
range of 0-8 for each of the four questions, for a total score range of 0-32. For each item,
the scoring followed the following basic pattern:

- 0 = no response or a response inappropriate to the task demand
- 2 = a vague or inaccurate response, or a response reversing the key metaphor of the
  selection
- 4 = accurate but limited response; does not address all three elements of the central
  metaphor (topic, vehicle, ground)
- 6 = response addresses all three elements of central metaphor correctly (meets
  expectations)
• 8 = response that addresses all three elements of central metaphor correctly and is also fluent (several sentences) in explanation (exceeds expectations)

The two forms of the instrument had been included in the previous pilot of the intervention and were revised based on teacher feedback. The revised versions were then piloted with a relevant group of children. As a preliminary step, nine second grade students tried out both forms of the test and were told to ask for clarification on any items they found confusing. Based on this tryout phase, the wording of two of the questions was revised for clarity. Then three other groups of second grade students participated in the next phase of the pilot, with two of the groups serving as pilot groups for test-retest reliability on the A form and the B form, respectively, and the third group serving to determine equivalent-forms reliability. Pilot tests were then scored independently by the researcher and another scorer with previous experience in using the original version of the rubric. The independent scoring process resulted in interrater reliability of .90 for the A version of the test and .88 for the B version, and subsequent discussion of conflicting scores resulted in 100% score agreement.

Equivalent forms reliability.

The pilot group to determine equivalent-forms reliability consisted of 11 second grade students, ranging in age from 7 years 4 months to 8 years 6 months with a mean age of 7 years 11 months. The group included six boys and five girls who took the test in a classroom setting. To account for potential fatigue and practice effects, the tests were administered in different order; six students took Form A followed by Form B, while the remaining five took Form B followed by Form A. The correlation coefficient for this
analysis was .507, which was not statistically significant ($p > .05$). Table 9 gives the means and standard deviations of both forms.

Table 9

**Means of LAIT Form Scores**

<table>
<thead>
<tr>
<th>Form</th>
<th>$M$</th>
<th>$SD$</th>
<th>$n$</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form A</td>
<td>9.09</td>
<td>2.59</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Form B</td>
<td>9.64</td>
<td>3.07</td>
<td>11</td>
<td>.507</td>
</tr>
</tbody>
</table>

To attempt to account for the nonsignificant correlation between the forms, student scores were examined individually. No patterns of test order or gender were evident. Differences between A scores and B scores ranged from 0 to 6, with the following additional breakdown: four students had identical scores on the two tests; three students had a difference of 2 points, with three having higher A scores and one higher B score; and two had a difference of 4 points, with one favoring A and one favoring B although both took the tests in B-A order. Only one student had a score difference of 6, with a higher score on the B test, which was taken second. Recognizing the powerful effect individual scores can have on results with a very small sample size, an additional analysis was conducted, excluding this student as an outlier. This second analysis yielded a statistically significant correlation of .660 ($p < .05$), with means and standard deviations as demonstrated in Table 10.
Table 10

Means of LAIT Form Scores, Outlier Removed

<table>
<thead>
<tr>
<th>Form</th>
<th>M</th>
<th>SD</th>
<th>n</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form A</td>
<td>9.20</td>
<td>2.86</td>
<td>10</td>
<td>.660*</td>
</tr>
<tr>
<td>Form B</td>
<td>9.20</td>
<td>2.70</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05

Test-retest reliability.

The test-retest reliability method was also utilized to assess the stability of the instrument (Janda, 1998). The two forms of the instrument were assessed separately for this purpose, with two different groups of students, each taking one form of the test twice with a three-week interval between administrations. The test-retest pilot group for this test was the same set of students as the pilot group for the comprehension test, with the same students taking the A forms of both tests or the B forms of both tests. Each group included 10 students.

The Pearson $r$ correlation coefficient showing the relationship between the two administrations of Form A was .844 ($p < .01$). The details of this analysis are shown in Table 11.

Table 11

Form A Test-Retest Reliability (LAIT)

<table>
<thead>
<tr>
<th>Administration</th>
<th>$M$</th>
<th>$SD$</th>
<th>$n$</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form A – first administration</td>
<td>10.60</td>
<td>3.13</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Form A – second administration</td>
<td>10.00</td>
<td>3.53</td>
<td>10</td>
<td>.844**</td>
</tr>
</tbody>
</table>

**$p < .01$**
The analysis of Form B also demonstrated a statistically significant correlation coefficient, with a value of .765 \((p < .01)\). The results for this analysis are shown in Table 12.

Table 12

*Form B Test-Retest Reliability (LAIT)*

<table>
<thead>
<tr>
<th>Administration</th>
<th>M</th>
<th>SD</th>
<th>n</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form B – first administration</td>
<td>8.60</td>
<td>3.53</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Form B – second administration</td>
<td>8.60</td>
<td>2.32</td>
<td>10</td>
<td>.765**</td>
</tr>
</tbody>
</table>

**\(p < .01\)**

*Teacher Assessment of Student Performance (TASP)*

A third instrument used in the study to assess students’ performance on tasks related to metaphor development throughout the course of the intervention was an assessment completed by the treatment group teachers for each child involved in the study. (See Appendix E.) The rationale behind the inclusion of this instrument was that much of the intervention involves discussion and whole-group product development under teacher guidance, neither of which could be assessed at an individual student level by the researcher, given the context of the study. Moreover, because it was not limited to performance on written tasks, the assessment could provide data on students whose paper and pencil performance did not reflect oral behaviors. Thus, the Teacher Assessment of Student Performance (TASP) was used to provide additional data on individual student performance based on the teachers’ involvement and observations.

The questions included in the instrument were aligned with the abilities given by Levorato (1993) as indicators of developing figurative competence. Six of the ten items
specifically reflect these abilities. The other items correspond with the abilities to identify topic, vehicle, and ground in a given metaphor assessed in the test of metaphor comprehension.

The TASP was reviewed by committee members, revised, and piloted with teachers for content validity and readability. Teachers then completed a form for each student at the conclusion of the intervention period and submitted forms with the post-assessments.

Classroom Observation Form (COF)

The fourth instrument included in the study was a classroom observation form, used to ensure fidelity of implementation of the unit and to mark differences among the group of treatment teachers. (See Appendix F.) This form includes items related to the three dimensions of the Integrated Curriculum Model and to the cognitive mediation around figurative competence incorporated within the intervention. The form is a revised version of a form in use with other units developed under the ICM; the original form was reviewed for clarification by a group of teachers and observers in the district involved in the proposed study and revised based on their commentary. In addition, many of the specific items on the form were drawn from a longer form on which content validity was given a mean rating of .96 by several experts in gifted and general education and on which interrater reliability analysis showed a median kappa of .63 (Avery, 1999). The revisions related to the metaphor study in particular were reviewed with teachers prior to the implementation of the intervention.
Study Intervention Procedures

The intervention tool of the study was an instructional unit, developed under the Integrated Curriculum Model (VanTassel-Baska, 1986, 1995), which embeds the study of metaphor within the context of a broader language arts curriculum framework that also incorporates writing instruction, literary analysis and interpretation, and vocabulary study. The unit framework and lesson outline are attached in Appendix G. Table 13 provides an outline of the specific lessons that emphasize metaphor comprehension and production and thus were the lessons required of the treatment group teachers.

Table 13

<table>
<thead>
<tr>
<th>Lesson</th>
<th>Metaphor Emphasis</th>
<th>Relevant Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
<td>• Informal discussion of metaphorical basis of Sandburg’s “Fog”</td>
<td>• Pre-assessment (LAIT) • Tree comparisons • Comparisons using clouds, trees</td>
</tr>
<tr>
<td>4. Adjectives</td>
<td>• Analysis of common characteristics of topic and vehicle in “Fog”; introduction of Metaphor Chart</td>
<td>• Selection of key vehicles for given adjectives</td>
</tr>
<tr>
<td>5. The Unicorn and the Moon</td>
<td>• Introduction to Literature Web</td>
<td>• Group-completed webs</td>
</tr>
<tr>
<td>7. Similes and Metaphors</td>
<td>• Introduction of terms simile and metaphor</td>
<td>• Metaphor Chart</td>
</tr>
<tr>
<td></td>
<td>• Additional emphasis on Metaphor Chart</td>
<td>• Moon poem analyses</td>
</tr>
<tr>
<td>8. <em>Owl Moon</em></td>
<td>• Composition of similes or metaphors about the sun</td>
<td>• Sun comparisons</td>
</tr>
<tr>
<td></td>
<td>• Analysis of story using Literature Web, Metaphor Chart</td>
<td>• Group webs and</td>
</tr>
<tr>
<td>10. Analogies</td>
<td>• Discussion of analogy as a form of metaphor</td>
<td>• Group webs and Metaphor Chart</td>
</tr>
<tr>
<td>11. Understanding Words in Context</td>
<td>• Analysis of poem demonstrating confusion of multiple meanings of words</td>
<td>• Student booklets demonstrating intentional confusion</td>
</tr>
<tr>
<td>12. Haiku and the Seasons</td>
<td>• Analysis of metaphors contained within haiku</td>
<td>• Student-developed haiku</td>
</tr>
<tr>
<td>13. Personification</td>
<td>• Analysis of story using Literature Web, Metaphor Chart</td>
<td>• Group webs and Metaphor Chart</td>
</tr>
<tr>
<td>14. Images and Symbols</td>
<td>• Analysis of poetry using Literature Web, Metaphor Chart</td>
<td>• Group and individual webs, Metaphor Chart</td>
</tr>
<tr>
<td>19. Unit Wrap-Up</td>
<td>• Final LAIT assessment</td>
<td>• LAIT posttest</td>
</tr>
</tbody>
</table>
The instructional unit makes use of two principal tools to promote students’ understanding of metaphor structure, the use of imagery in literature, and literary analysis and interpretation in general. One of these tools, the Metaphor Chart, is illustrated in Figure 2 with examples from the unit. This tool encourages analysis of the structure of metaphors found within texts and emphasizes the connections between topic and vehicle through the relevant ground. The other tool, illustrated in Figure 3, is a Literature Web that encourages analysis of literature through discussion of key words, feelings, central ideas, and important images and symbols. Within the context of the intervention, this tool was utilized to encourage students to note the important words and images, to analyze them as metaphors as appropriate, and to make connections between the metaphors and the ideas contained within text. These two tools, along with specific discussion questions and follow-up activities, formed the core instructional pieces of the intervention and also comprised a major part of the training workshop in which teachers participated prior to their work in the study.
### Metaphor Chart

<table>
<thead>
<tr>
<th>Topic of the Comparison</th>
<th>What it is Compared To</th>
<th>Important Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>fog</td>
<td>cat</td>
<td>quiet, gray, sneaky</td>
</tr>
<tr>
<td>moon</td>
<td>clock face</td>
<td>round, can look like a face</td>
</tr>
<tr>
<td>moon</td>
<td>cooky</td>
<td>round, gets smaller, gets</td>
</tr>
<tr>
<td></td>
<td></td>
<td>bigger (baking)</td>
</tr>
<tr>
<td>moon</td>
<td>coin</td>
<td>shiny, round</td>
</tr>
<tr>
<td>trees</td>
<td>giant statues</td>
<td>tall, stand still</td>
</tr>
<tr>
<td>daisies in grass</td>
<td>stars in sky</td>
<td>bright, white, shape</td>
</tr>
<tr>
<td>night</td>
<td>train</td>
<td>moves, carries things,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>interesting and mysterious</td>
</tr>
</tbody>
</table>

*Figure 2.* Metaphor Chart employed in instructional unit. Headings correspond to metaphoric elements of topic, vehicle, and ground. Sample responses drawn from literature selections in lessons 1, 7, 8, 10, and 14 of unit.
Figure 3. Literature Web employed in unit. This is a slight modification of a web used in other language arts units for older students.

The unit was piloted twice by the researcher within the context of a university-based enrichment program with gifted students in grades K-2 and revised based on the experiences with these pilots. The unit was also piloted by three teachers within the district that served as the study site. Two of these teachers piloted the unit with second graders designated as advanced readers by the district reading test discussed previously; the third teacher piloted the unit with a group of identified highly gifted second graders. Commentary from these teachers was also incorporated into the revision process. In addition, the unit was reviewed by an expert in curriculum development for the gifted in language arts and revised based on her commentary. The version used in the study was reviewed by an expert in early childhood education and metaphor development for content validity, and this expert confirmed that the unit did indeed provide direction for teaching the structure and comprehension of metaphor.

Teachers to be involved in the study were introduced to the unit through a two-day training workshop given by the researcher and a teacher trainer in the school district, who had also served as one of the pilot teachers for the unit, with additional assistance from the other two pilot teachers. This training workshop addressed many elements of the overall unit, including the ICM under which it was designed and the specific teaching models it incorporates. Heavy emphasis during the workshop was placed on literary analysis and interpretation, including modeling of the use of the Literature Web, and on the definitional structure of metaphor as discussed in Chapter II, including modeling of the use of the Metaphor Chart.

The teachers in the treatment group were given a list of lessons required for implementation and a list of student products to collect during the period. In addition, the
teachers were given copies of the TASP (Appendix E) to complete for each student in the study group.

Implementation of the intervention took place between mid-February and early May, 2001. Teachers were asked to administer the pre-tests before beginning unit lessons and to administer the post-tests after completion at least through lesson 14. Comparison groups were undergoing literature-based language arts instruction during this period as well, although comparison teachers did not have access during the period of implementation to the intervention materials. Instruction in comparison classes during the second grade year, prior to or at the same time as treatment intervention also included some emphasis on poetry and identification of the figurative language devices of metaphor and simile, but without the specific instructional models utilized in the study intervention.

**Data Collection**

Tests were administered by the teachers involved in the study, with instructions for administration given by the researcher. The test of metaphor comprehension was scored based on the protocol given in Appendix B. The test of literary analysis and interpretation was scored by the researcher and a second trained scorer, using the rubric described in the instrumentation section above and included in Appendix D. Interrater reliability with this rubric was .90. In addition, responses to Question 2 on this test were analyzed for level of metaphor interpretation based on the semantic mapping framework outlined by Johnson and Pascual-Leone (1989).

Student products were collected from some teachers relating to lessons of the unit as outlined in Table 13 for qualitative analysis. The qualitative portion of the study was
conceptualized as a multi-site case study (Creswell, 1998), with the “bounded system” under exploration the implementation of the intervention itself. The products and anecdotal data, then, representing both individual efforts and group work from the students, were analyzed across the treatment group for themes and patterns (Yin, 1994).

During the implementation period, the researcher conducted two observations in four of the five treatment classrooms and one observation in the fifth. These observations served two purposes: first, to ensure fidelity of implementation of the unit lessons, and second, to gather anecdotal data on student response to the unit. Comparison group teachers administered the pretests in late February and the posttests in early May.

**Data Analysis**

Several methods of data analysis were used to determine the results of this study. The primary method was an analysis of covariance to determine any differences between performance of the treatment and comparison groups on the two posttests, with the pretest scores for both groups serving as the covariate (Gall, Borg, & Gall, 1996). Correlational analyses were then conducted to determine relationships of student performance on the tests to their scores on the district’s screening measures for gifted and talented identification, the Raven Standard Progressive Matrices and the Test of Cognitive Skills. Correlations were also determined between performance on the written tests and performance as assessed by the teachers on the teacher assessment form.

A content analysis was conducted on the student products and anecdotal records, representing the qualitative portion of the data analysis process (Creswell, 1994, 1998). Patterns of student response were established, drawn from the literature relating to the stages of metaphor development (Broderick, 1991; Johnson & Pascual-Leone, 1989;
Levorato, 1993; Vygotsky, 1987), and then products and observations were examined for
demonstrations of the established patterns or of additional patterns (Yin, 1989).

Classroom observation form data were analyzed descriptively to report fidelity of
implementation of the intervention. These data were also analyzed with student
performance data to determine existing differences between changes in performance with
the treatment group based on teacher performance differences.

Sub-analyses were also conducted across instruments to explore any performance
differences based on age or gender.

*Delimitations and Limitations*

The delimitations placed on the study by the researcher included several issues
related to sampling and instrumentation. The sample was limited to one grade level, a
limited age range, and a limited range of demonstrated reading achievement levels. This
decision was made in order to give more control over unit implementation to the
researcher and to limit the confounding effect of student reading level on text-based tests
of figurative competence. The intervention occurred in the natural setting of second grade
classrooms with intact groups of students, so as to measure the effects of the intervention
itself in a non-laboratory environment.

The combination of two paper-and-pencil instruments assessing metaphor
comprehension was used in order to accommodate the problem of asking students to
respond in a linguistically complex format (Vosniadou, 1987) but also to provide
opportunity for a greater range of responses and to limit the effects of chance selection of
correct responses. The decision to focus the tests on metaphor comprehension rather than
production is based on the conceptual framework and Levorato’s (1993) Model of
Figurative Competence: according to this model, children in the age range of students in the study are developing the abilities related to comprehension more readily than those related to production. However, with the assumption in mind that some students may demonstrate levels of development beyond the first three, demonstration of production of metaphor was examined within student portfolios and teacher assessments.

Several limitations existed within the sampling, instrumentation, and procedures of the study. The use of intact groups limits generalizability, as does the implementation of the study within a single school district. Furthermore, comparisons were made across classes of students, even though they were taught by different teachers in different classrooms and schools. Moreover, this aspect is a limitation because of the differences in teacher performance during the unit, including both specific demonstration of key instructional elements during observed classes and differences in the numbers of lessons implemented. The issue of volunteerism also adds a limitation to the study; all teachers involved volunteered to participate, which may reflect certain characteristics in the personalities and/or teaching styles of these teachers that may not be generalizable to the larger teaching population.

The comparison group presented additional limitations. The comparison group was considerably smaller than the treatment group, and it was not drawn from the same schools as the treatment classes. One of the two comparison classes came from the same school district, but a second comparison class had to be drawn from a different district because no additional comparison teachers came forward in the district in which the study was taking place.
Generalizability was also limited by the nature of the intervention used. A specific instructional unit of several lessons was used as the intervention, so results may be closely related to that specific intervention rather than to instruction in metaphor development in general.

Another limitation exists because of the use of several researcher-designed instruments. The intervention was designed and revised by the researcher, although piloted by other teachers as well. The test of metaphor comprehension and the teacher assessment were also designed by the researcher, based on the review of the literature for optimal test conditions related to metaphor comprehension. Moreover, the literary analysis and interpretation test and rubric and the classroom observation form were pre-existing but were tailored to this study by the researcher. All instruments were reviewed and piloted; however, small pilot sample sizes and low to nonsignificant reliability coefficients on some aspects of the instruments make inferences about results cautionary.

**Ethical Considerations**

The study was conducted in a manner designed to protect the anonymity of the subjects involved and to ensure that all of the participants could benefit from any favorable outcomes. The study proposal was approved by Human Subjects Review at the College before any testing was conducted. Permission from school district administration and building principals for the instrument pilot and the study was sought and given, and parents and students involved signed a permission form indicating their understanding of study parameters and the voluntary nature of involvement. Student data were maintained based on number, and all results were reported anonymously.
Conclusion

The foregoing pages have outlined the procedures and instruments used to gather data in this study of metaphor development. The following chapter will address these issues further as it presents findings related to each of the research questions, drawn from the study instruments described above.
CHAPTER IV

Findings

Introduction

This chapter summarizes findings related to each of the four questions exploring metaphor development in second grade students and the effects of an instructional intervention on their demonstrations of metaphor comprehension and production. The study included a treatment group and a comparison group, as outlined in Chapter III, and each group was tested before and after the intervention using two measures of figurative competence. Moreover, additional data were collected from the treatment group: at least one classroom observation was conducted in each of the five classes, and student products and teacher assessments of students’ figurative competence were also collected. The following sections address findings related to each of the data sources and research questions. First addressed are research Questions 1 and 2, which primarily focused on students’ demonstration of figurative competence on study instruments and products; Questions 3 and 4 then address findings related to the intervention and its effects.

Findings Related to Research Question 1

Research Question 1 addressed the issue of to what degree student performance on measures of figurative competence was related to their scores on other measures of verbal and general ability and achievement. This relationship was measured by comparing student gain scores and posttest scores on the two measures of figurative competence utilized in the study to their scores on standardized tests provided by the school district. The scores provided by the school district were the students’ scores on the measures used for gifted program identification, administered in the spring of the second
grade year unless previously requested by teacher or parent. Because these were the only
scores the district provided, they represent only ability scores and not achievement
scores. In the study group, all student scores used in the analyses that follow were results

Scores were provided on the Test of Cognitive Skills (TCS) for four of the five
treatment classes and one comparison class. This test includes four subtests, addressing
analogies, memory, sequencing, and verbal reasoning. For the first three of these subtests,
a raw score of 20 represents a perfect score, while on the verbal reasoning test a score of
19 represents the ceiling. Raw scores for all four subtests were provided by the schools.
For each subtest, a $t$-test comparing the results for the two groups was performed with no
statistically significant results, although any conclusions about the equivalence of the
groups must be made with caution because of the small size of the comparison group.

Results of the $t$-tests are given in Table 14.

Table 14

*Comparison of TCS Subtest Results for Treatment and Comparison Groups*

<table>
<thead>
<tr>
<th>Test</th>
<th>Treatment Group</th>
<th></th>
<th>Comparison Group</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Verbal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reasoning</td>
<td>16.28</td>
<td>1.96</td>
<td>15.85</td>
<td>2.41</td>
</tr>
<tr>
<td>Analogies</td>
<td>17.11</td>
<td>3.06</td>
<td>17.46</td>
<td>2.30</td>
</tr>
<tr>
<td>Sequencing</td>
<td>18.14</td>
<td>2.80</td>
<td>18.00</td>
<td>1.91</td>
</tr>
<tr>
<td>Memory</td>
<td>12.47</td>
<td>4.99</td>
<td>10.69</td>
<td>4.35</td>
</tr>
</tbody>
</table>
The descriptive results given in Table 14 demonstrate that on several of the subtests, notably the analogy and sequencing subtests, the students in both the treatment and comparison groups achieved mean scores that neared the ceiling for the tests. Indeed, for each subtest, at least ten percent of the sample achieved a perfect score, and on the sequencing subtest a full 41% achieved a perfect score. The high mean scores of both groups and limited variability affect the degree to which strong predictive relationships can be determined between these tests and the instruments utilized in the study, but analyses were conducted nevertheless to explore any significant findings.

The results of the TCS were analyzed with posttest scores and with gain scores on both study instruments across the treatment and comparison groups. Posttest scores and gain scores were used in separate analyses in order to assess existing relationships both between performance on two verbal tests at approximately the same time (the posttest and TCS) and between a verbal ability test and measured growth, reflecting potential for development growth gains. Treatment and comparison group scores were all included in the analysis, with grouping entered as a possible predictor. Descriptive results on the posttest and gain scores on the MCT are given in Table 15.
Table 15

*Descriptive Results on MCT*

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Gain Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>15.56</td>
<td>14.36</td>
<td>-1.01</td>
</tr>
<tr>
<td>SD</td>
<td>3.95</td>
<td>4.09</td>
<td>3.77</td>
</tr>
<tr>
<td>n</td>
<td>35</td>
<td>69</td>
<td>35</td>
</tr>
<tr>
<td><strong>Comparison</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>13.42</td>
<td>13.18</td>
<td>-.24</td>
</tr>
<tr>
<td>SD</td>
<td>4.49</td>
<td>4.31</td>
<td>3.53</td>
</tr>
<tr>
<td>n</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>14.75</td>
<td>14.08</td>
<td>-.72</td>
</tr>
<tr>
<td>SD</td>
<td>4.25</td>
<td>4.15</td>
<td>3.67</td>
</tr>
<tr>
<td>n</td>
<td>56</td>
<td>90</td>
<td>56</td>
</tr>
</tbody>
</table>

As with the TCS subtests, these MCT posttest results also indicate a possible ceiling effect. Within both the treatment and comparison groups, a higher percentage of students scored above the researcher-determined 85% cut-off than had done so in the pilot group. On the pretest, six students (17.1%) in the treatment group and three students (14.2%) in the comparison group exceeded 85% correct, with one student in the treatment group achieving a perfect score. On the posttest, twelve students in the treatment group (17.3%) and two students in the comparison group (9.5%) exceeded 85% correct. All of
these calculations exceeding or nearing 10% of the sample suggest a possibility of ceiling effect with the instrument, particularly the treatment group results nearing 20% on both forms. This issue and the troubling negative gain scores for both groups are addressed in more detail in the discussion of research Question 3.

A multiple regression analysis was performed to determine existing significant relationships between the TCS subtests and the instruments used in the study (Grimm & Yarnold, 1995). The first analysis used gain scores on the MCT as the dependent variable and entered the four subtests of the TCS to determine significant relationships. This analysis yielded no relationships sufficiently strong to develop a predictive model, as shown in Table 16.

Table 16

<table>
<thead>
<tr>
<th>Subtest</th>
<th>B</th>
<th>SEB</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Reasoning</td>
<td>.43</td>
<td>.34</td>
<td>.25</td>
</tr>
<tr>
<td>Analogies</td>
<td>-.12</td>
<td>.20</td>
<td>-.12</td>
</tr>
<tr>
<td>Sequencing</td>
<td>-.30</td>
<td>.26</td>
<td>-.24</td>
</tr>
<tr>
<td>Memory</td>
<td>-.01</td>
<td>.14</td>
<td>-.14</td>
</tr>
</tbody>
</table>

*Note. R² = .11 (n = 36)*

A second regression analysis was conducted on the MCT, this time utilizing posttest scores instead of gain scores as the dependent variable. This analysis yielded a statistically significant correlation between the MCT posttest scores and the TCS verbal reasoning subtest and a significant model, although with a low R² value. Table 17 displays the results of this analysis.
Table 17

*Regression Analysis Summary for TCS Subtests Predicting MCT Posttest Scores*

<table>
<thead>
<tr>
<th>Subtest</th>
<th>$B$</th>
<th>$SE_B$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Reasoning</td>
<td>.98</td>
<td>.29</td>
<td>.46**</td>
</tr>
<tr>
<td>Analogies</td>
<td>.11</td>
<td>.19</td>
<td>.09</td>
</tr>
<tr>
<td>Sequencing</td>
<td>-.15</td>
<td>.22</td>
<td>-.10</td>
</tr>
<tr>
<td>Memory</td>
<td>.21</td>
<td>.14</td>
<td>.25</td>
</tr>
</tbody>
</table>

*Note. $R^2 = .27$ (n = 47, $p < .05$)*

**$p < .01$**

Similar analyses to those discussed above for the MCT were also conducted using results from the LAIT. Descriptive results for the LAIT for the treatment and comparison groups are given in Table 18.

Table 18

*Descriptive Results on LAIT*

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Gain Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>9.27</td>
<td>11.42</td>
<td>2.07</td>
</tr>
<tr>
<td>$SD$</td>
<td>2.38</td>
<td>2.38</td>
<td>3.03</td>
</tr>
<tr>
<td>$n$</td>
<td>55</td>
<td>59</td>
<td>55</td>
</tr>
<tr>
<td>Comparison</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>10.67</td>
<td>10.78</td>
<td>.11</td>
</tr>
<tr>
<td>$SD$</td>
<td>3.07</td>
<td>2.67</td>
<td>4.68</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
The LAIT results did not suggest the ceiling effect issues indicated by MCT results; no students in the sample scored near the possible total score of 24. As with the MCT, multiple regression analyses were conducted to determine relationships between LAIT scores and the subtests of the TCS. An analysis using gain scores on the LAIT as the dependent variable yielded a predictive model demonstrating a small but significant correlation between gain scores and the verbal reasoning subtest; however, this correlation was actually a small negative correlation. The results of this analysis are displayed in Table 19.

Table 19

<table>
<thead>
<tr>
<th>Subtest</th>
<th>B</th>
<th>SEB</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Reasoning</td>
<td>-.67</td>
<td>.25</td>
<td>-.36**</td>
</tr>
<tr>
<td>Analogies</td>
<td>.01</td>
<td>.16</td>
<td>.08</td>
</tr>
<tr>
<td>Sequencing</td>
<td>-.23</td>
<td>.19</td>
<td>-.16</td>
</tr>
<tr>
<td>Memory</td>
<td>.15</td>
<td>.10</td>
<td>.21</td>
</tr>
</tbody>
</table>

Note. $R^2 = .17$ ($n = 57, p < .05$)

**$p < .01$
An analysis using LAIT posttest scores as the dependent variable instead of gain scores yielded a significant correlation between posttest scores and the memory subtest, but the model overall did not yield a statistically significant result. The details of this analysis are displayed in Table 20.

Table 20

*Regression Analysis Summary for TCS Subtests Predicting LAIT Posttest Scores*

<table>
<thead>
<tr>
<th>Subtest</th>
<th>B</th>
<th>SEB</th>
<th>( \beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Reasoning</td>
<td>-0.31</td>
<td>0.18</td>
<td>-0.22</td>
</tr>
<tr>
<td>Analogies</td>
<td>-0.00</td>
<td>0.12</td>
<td>-0.04</td>
</tr>
<tr>
<td>Sequencing</td>
<td>-0.01</td>
<td>0.14</td>
<td>-0.09</td>
</tr>
<tr>
<td>Memory</td>
<td>0.16</td>
<td>0.07</td>
<td>0.31*</td>
</tr>
</tbody>
</table>

*Note. \( R^2 = 0.14 \) (n = 58)*

*p < .05*

*Summary of Findings Related to Research Question 1*

Regression analyses intended to assess relationships between study instruments and other measures of verbal and general ability yielded few significant results, with only small correlations existing between students' scores on the study instruments and their scores on the verbal reasoning subtest and in one case the memory subtest of the Test of Cognitive Skills. Models reflecting statistically significant correlations with the verbal reasoning subtest emerged with posttest scores on the MCT and gain scores on the LAIT, although in the latter the predictive relationship was a negative one. A small but significant correlation between LAIT posttest scores and memory subtest scores was insufficient to yield a predictive model. Limited variability within the sample and
clustering of scores near a low ceiling on the TCS subtests likely affected the possibility of determining statistically significant relationships.

Findings Related to Research Question 2

Research Question 2 was intended to investigate a broader selection of student performance items than could be assessed solely through use of the two measures of figurative competence. Student products were collected from all treatment classes and anecdotal records were kept during classroom observations in order to provide illustrative data regarding students' comprehension and production of metaphor. These products were collected, catalogued, and then assessed within the categories of comprehension and production for evidence of patterns reflecting prior research findings on metaphor development in children, in accordance with Yin's (1994) recommendations for pattern-finding methods in qualitative data analysis.

Metaphor Production

Collected student products and records of classroom observations were reviewed for instances in which students provided written statements that included comparisons of two different items. In most cases, the tasks from which these records were drawn were presented to students as metaphor production tasks, although in some instances they were merely asked to write statements that compared two items. These comparative statements from students were then reviewed according to five patterns previously studied in the research on metaphor development. First, statements were examined for evidence of metaphorical intent through clear violation of conventional categories in the comparison (Vosniadou, 1987). Statements that were not rated as violating conventional categories were not analyzed further, as they were not to be considered metaphorical. Second, the
statements were examined on the basis of appropriateness of the metaphor, rated as appropriate, flawed, or vague (Broderick, 1991). Third, the statements were examined for the basis of the comparison, whether they were attributional or relational metaphors (Gentner, 1988). Fourth, the statements were assessed on the categorical dimension into which the ground of the metaphor fell (Broderick, 1991). Fifth, the statements were assessed for the relative degree of salience of the topic and vehicle terms (Glucksberg & Keysar, 1993). Beyond these five specific qualities of the comparisons, the statements were reviewed for evidence of other interesting or unusual features, noted as appropriate throughout the discussion below.

A total of 47 statements were reviewed according to these specific emphases. Among the statements were six that included two separate comparisons, leading to a total number of 53 assessed comparisons. The statements were drawn from three of the five treatment classes and included samples from four different activities from the unit. Table 16 illustrates the breakdown of the statements collected and the activities from which they were drawn.

Table 21

Comparative Statements from Treatment Classes

<table>
<thead>
<tr>
<th>Activity</th>
<th>Number of Statements</th>
<th>Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjectival similes – Students were asked to create a comparison using a common adjective and the structure &quot;____ is as ____ as a ____.&quot;</td>
<td>12</td>
<td>1 (Class 4)</td>
</tr>
</tbody>
</table>
Haiku – Students were asked to choose something in nature and write a haiku response, using a metaphorical comparison.

Tree comparisons – Students were asked to write short “poems” comparing a tree to something it was like.

Sun comparisons – Students were asked to write similes or poems comparing the sun to something it was like.

Note. Class numbers refer to numbers given in Table 3 in Chapter III.

The sorting and classification of the students' comparative statements are discussed in the sections that follow and summarized in Table 22. A chart listing the specific statements and demonstrating graphically the analyses discussed is included in Appendix H. In the discussion and in the chart in Appendix H, students' spelling is reproduced as originally written.

Table 22

Quantitative Summaries of Comparative Statements

<table>
<thead>
<tr>
<th>Classification</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classifiable as Metaphor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category violation – metaphorical</td>
<td>45</td>
<td>84.9</td>
</tr>
<tr>
<td>No category violation – non-metaphorical</td>
<td>8</td>
<td>15.1</td>
</tr>
<tr>
<td>Appropriateness</td>
<td>31</td>
<td>68.9</td>
</tr>
<tr>
<td>------------------------</td>
<td>----</td>
<td>------</td>
</tr>
<tr>
<td>Appropriate</td>
<td>31</td>
<td>68.9</td>
</tr>
<tr>
<td>Flawed</td>
<td>12</td>
<td>26.7</td>
</tr>
<tr>
<td>Vague</td>
<td>2</td>
<td>4.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type of Comparison</th>
<th>32</th>
<th>68.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attributional</td>
<td>32</td>
<td>68.9</td>
</tr>
<tr>
<td>Relational</td>
<td>7</td>
<td>15.6</td>
</tr>
<tr>
<td>Both</td>
<td>5</td>
<td>13.3</td>
</tr>
<tr>
<td>Unclassifiable</td>
<td>1</td>
<td>2.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category of Comparison</th>
<th>24</th>
<th>53.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual</td>
<td>24</td>
<td>53.3</td>
</tr>
<tr>
<td>Tactile</td>
<td>5</td>
<td>11.1</td>
</tr>
<tr>
<td>Visual/Tactile Combination</td>
<td>2</td>
<td>4.4</td>
</tr>
<tr>
<td>Sound</td>
<td>1</td>
<td>2.2</td>
</tr>
<tr>
<td>Behavior/Function</td>
<td>7</td>
<td>15.6</td>
</tr>
<tr>
<td>Natural Cycles/Time</td>
<td>5</td>
<td>11.1</td>
</tr>
<tr>
<td>Unclassifiable</td>
<td>1</td>
<td>2.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Salience</th>
<th>18</th>
<th>40.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Salience Topic – High Salience Vehicle</td>
<td>18</td>
<td>40.0</td>
</tr>
<tr>
<td>Similar Salience in Topic and Vehicle</td>
<td>22</td>
<td>48.9</td>
</tr>
<tr>
<td>High Salience Topic – Low Salience Vehicle</td>
<td>3</td>
<td>6.7</td>
</tr>
<tr>
<td>Unclassifiable</td>
<td>2</td>
<td>4.4</td>
</tr>
</tbody>
</table>
Note. Percentages related to Appropriateness, Type, and Category of Comparison as well as Salience are from a total \( n \) of 45 metaphorical statements.

Metaphorical intent.

Of the 53 statements assessed, 45 demonstrated violation of conventional categories in the comparisons they offered. In two of the four activities, the sun comparisons and the haiku activity, students were specifically asked to write metaphors, so metaphorical statements could be expected from these specific activities. Three of the eight non-metaphors occurred in the sun comparison activity, while three occurred in the tree comparison and two in the adjectival similes. In three of the eight cases, the comparison made was definitively literal as opposed to metaphorical: one student commented that “Mr. Z’s shirt is as green as Cody’s shirt,” while another wrote that “A tree is like a tall plant,” and a third wrote that “The sun is like a star.” This latter statement, however, although this portion was rated as non-metaphorical, also included a further sentence that held a metaphorical comparison: “The sun is like a star. Big bright and butiful waiting all day for moon to rise so he can go to sleep.” This statement contains a literal comparison (sun = star) and an appropriate metaphorical one (sun = something that sleeps). An additional two statements were rated as non-metaphorical because they offered items from different categories but discussed how the items differed rather than how they were alike; one child explained how a tree was different from broccoli, while another discussed how trees and telephone poles differed.

Recalling Addison’s (1993) and Vosniadou’s (1987) notions of a continuum from literal to figurative, the remaining three non-metaphorical statements could be placed further toward the figurative on the continuum, although still not sufficiently
metaphorical to be rated as violating conventional categories. One student wrote that “Water can be as cold as ice,” while another wrote that “The sun is a white hot fire slowly burning itself out,” both of which provide vehicles that are too close in nature to the topics to be considered metaphorical. The final statement, “The sun is a ball of fire circling around in space,” was more nearly metaphorical, but it is not clear whether the term “ball” is used as a descriptor of the spherical shape or as a deliberate comparison to a ball such as might be used in a game; thus, the statement was rated non-metaphorical.

Once the statements had all been rated as metaphorical or non-metaphorical, the further analyses focused only on the 45 statements judged to be metaphorical. Each of these was assessed based on the appropriateness of the comparison, the type of comparison, and additional details related to the type and salience of the ground.

**Appropriateness of metaphorical comparison.**

Broderick (1991) found that as children progressed through the preschool and early primary years, their use of appropriate similes as opposed to flawed or vague ones increased, as did their recognition of appropriate similes. In this study, the 45 statements rated as metaphorical were examined and classified as appropriate, flawed, or vague, using Broderick’s system of two-step review: first, the statements were read for appropriateness and classified as appropriate if “acceptable or good” and as inappropriate if “awkward or unacceptable.” The inappropriate statements were then reviewed again for the salience of the ground characteristic in the vehicle term. Those inappropriate statements in which the salience was low were rated as vague, while those with high salience were rated as flawed.
Across the 45 statements provided by students in this study, only two were classified as vague, reflecting Broderick’s finding that vague comparisons disappear from students’ production earlier than flawed. Both of these statements were drawn from the tree comparison activity, which was conducted with Class 2 relatively early in the school year (October). The two vague comparisons related trees to “monsters” and to “a king” or “a bell that doesn’t ring,” respectively. The former statement, when elaborated, was clearly related to the writer’s feeling of a measure of fear of trees, while the latter statement appeared to place a priority on rhyme as opposed to appropriateness of the comparison.

Of the remaining 43 statements, 12 were rated as flawed because of their awkwardness or the weak nature of the comparison. Most of these statements were drawn from the tree comparison activity and the adjectival simile activity; only two were drawn from the sun comparison activity. Several of the statements from the adjectival simile activity, in which students had to provide topic, vehicle, and ground, were flawed because of the choice of a topic already very high in salience. For example, one student wrote that “my blood is as red as a red crayon,” using a high-salience topic and a vehicle that required an adjective to make it fit into the comparison in the first place. Another student wrote that “a ball is as round as the moon,” again providing a high salience topic. Other statements were flawed because they offered weak pairings of a topic and vehicle that shared a ground but in such different ways as to make the metaphor incongruent: “The sun is like the humongous Pacific Ocean swaying back and forth in space”; “The sun is as huge as 95,000,000 World Trade Centers glued together like an odd rubber ball”; “Trees are very big, they are big like a pig.” A third group of flawed statements
were characterized by the relatively even salience of topic and vehicle and the limited
degree to which the comparison offered additional information about the topic: “The
house was as big as an elephant”; “The baby was as small as a teddy bear”; “A tree is like
a streetlight.”

All of the remaining comparisons were rated as appropriate, because they offered
acceptable vehicles for comparison and expressed them in a sufficiently clear way as to
emphasize the intended ground. As previously mentioned, one submission included a
non-metaphorical comparison and a metaphorical one; another submission included both
a flawed and an appropriate metaphor: “Trees are like a pillow canectid to a sick....Trees
are a playground and a home at the same time.” This statement includes a flawed
metaphor and an appropriate one. In addition to these, four of the collected products
included two appropriate metaphors within the same statement; these specific products
will be discussed shortly. There was a total, then, of 31 appropriate comparisons, with
statements containing two counted twice.

Type and category of metaphorical relationship.

All of the metaphorical statements, including appropriate, flawed, and vague
statements, were assessed for the type of relationship they illustrated – whether the
metaphor was attributional, based on physical or sensory descriptors, or relational, based
on function, behavior, or analogical relationships (Gentner, 1988), with an expected
pattern of greater use of attributional than relational metaphors. Thirty-two of the
metaphors in this sample were rated as attributional, with seven relational metaphors and
five that were both attributional and relational. One metaphor was unclassifiable into
these groups because its vagueness made assessment of the intended ground impossible. This latter statement was the comparison of a tree to a king or bell discussed above.

Among the 32 attributional metaphors in the sample, the majority utilized visual attributes of the topic and vehicle to create comparison, with the most common types of attributes being color, shape, and size. Color comparisons included “the classroom was as colorful as a rainbow” and several that compared the sun to other things yellow or gold: “the sun is like a peice gold but you never touch or hold”; “the sun is like a lemon drop”, and “the sun is a hot yellowish light bulb.” Each of these comparisons also emphasized another visual or tactile attribute – shape or temperature. Shape was another commonly emphasized area for the metaphorical ground, with the most common vehicle being a ball. The topics of the sun, the moon, and rocks were all compared to balls in student metaphors, some with additional elaboration: “Rocks can be cool things/they’re all different sizes/just like lumpy balls”; “the sun is a golden ball shimmering in the sky.”

The third common visual attribute utilized was size, again frequently in conjunction with the attribute of shape within the same statement: “The sun is like a giant light bulb.” Several of the size-based metaphors were among the flawed metaphors discussed above, because they utilized two things expected to reflect the same size attributes and thus did not emphasize clearly details about the topic: e.g., “the house was as big as an elephant”; “the baby was as small as a teddy bear.”

After visual attributes, the next most common type of attributional comparison was a comparison based on tactile characteristics. Several metaphors emphasizing texture or temperature appeared in the sample. The texture-based metaphors, notably, were drawn from the activity in which the ground was given to students and they were to
create topic and vehicle: “my cat is as soft as a pillow” and “the football player was as rough as a rock.” The temperature metaphors all appeared in the sun comparisons, including “the sun is like a burning stove,” “the sun is like a bubbling, bristling bowl of hot water,” and several metaphors that made reference to fire or to heat within other comparisons.

One attributional metaphor used sound as the ground for comparison, but again this appeared in the activity in which students were given the ground as part of the task. Moreover, this comparison reflected a metaphor used commonly enough to reflect cultural/linguistic knowledge rather than metaphorical thinking: “the children were as quiet as a mouse.”

The seven relational metaphors reflected characteristics of behavior, function, or cycles in nature, generally demonstrating more complexity and abstraction than the attributional metaphors. For example, one student wrote that “the teacher roars as much as a lion,” emphasizing behavior, while another wrote that “Trees are a playground and a home at the same time,” illustrating function. Another child assigned behavioral characteristics to a mountain in a haiku that also included an attributional metaphor comparing snow to sugar: “A soft mountain so tall/reaching to grab the sugar/snow to be so white.” Two children used the sun and the moon in metaphors that related the natural cycles of these bodies to animal behavior: “The sun is first it stays until the moon takes birth”; “The sun is like a star. Big bright and butiful waiting all day for moon to rise so he can go to sleep.” One child created a complex metaphor in which one relational comparison was elaborated with a second relational comparison: “The sun is like a ball of sring, weving it’s lihght through space like a never-ending pecie of thred. Nether old,
nether yong, it's light is an everlasting memory.” This statement, which at first appears to be attributional in nature because of its comparison of shape, adds a level of abstraction when the function of weaving is incorporated. The comparison of light to memory and the allusion to age in the second portion represents another relational metaphor, again referring to time in nature, similar to the cyclical metaphors noted above. This last example verges upon the final group of metaphors assessed, which were classified as both attributional and relational at once.

Five metaphors were classified as both attributional and relational. All of these were drawn from the sun comparison activity. Two of the five metaphors again utilized the cycle of day and night for their comparison but included more attributional elaboration than the cyclical comparisons mentioned above: “The sun is a golden ball shimmering in the sky…. The stars play with throwing it up really high. Then it goes to the other side and the moon gets thrown over again.” “The sun is like hot yellow sand that gets wet by waves, goes out, and turns into night.” A third metaphor emphasized the heat of the sun but alluded as well to the movement (behavior) of the sun by comparing it to flowing lava: “The sun… is so burning hot. It pours over the earth like lava. Like so much burning fire.” The fourth metaphor combining attributional and relational grounds was a very unusual comparison in the reasoning it offered: “The sun is like water, sometimes nice and bright. But sometimes hot and mean and draws famine in.” The final statement in this category relied heavily on visual attributes of the sun but used them in what amounted to a comparison to the behavioral characteristics of storms: “The sun is like a fiery yellow storm of yellowish rays of snow.”
These five metaphorical statements all made effective use of relational and attributional characteristics and were also notable in that they maintained the focus of the metaphor through the elaboration provided. In other cases in the sample in which students wrote more than a simple statement, they frequently provided a second, unrelated metaphor or descriptive elaboration that did not maintain the metaphorical emphasis.

Relative salience of topic and vehicle.

Glucksberg and Keysar (1993) and Ortony et al. (1984) discussed the power of metaphors in terms of the relative salience of the topic and vehicle terms. Because the intent of a metaphor is to provide additional information or point of emphasis about the topic through the comparison to the vehicle, a metaphor is more effective if the vehicle is high in salience in terms of the ground characteristic; indeed, the vehicle should be higher in salience than the topic in order for the emphasis to be clear. If a metaphor is characterized by a topic higher in salience on the ground characteristic than the vehicle, then the vehicle is not useful in emphasizing the presence of the ground in the topic. Metaphors in which the topic and vehicle are approximately even in their salience in terms of the ground are also less effective than those with a low-high salience pattern. However, this is an advanced level of metaphorical understanding, and it was not expected that students in this study would necessarily be aware of the need for higher salience vehicles than topics. Research from other studies indicates that a pattern of preference for metaphors that are effectively asymmetric in terms of salience does not develop until the late elementary years (Glucksberg & Keysar, 1993; Kogan, Chadrow, & Harbour, 1989). Indeed, because many of the collected products related to the sun, a topic with very high salience of its key characteristics of light, heat, and size, it was expected
that no clear pattern of low-high salience patterns would emerge from the sample. However, salience is related to appropriateness of the metaphor, and thus the prediction that students would demonstrate fewer vague and flawed metaphors than appropriate ones also suggested a prediction that fewer high-low salience patterns would emerge than low-high or even patterns.

The metaphorical statements developed by the students in this study were rated according to the relative degree of salience of the ground existing in the topic and vehicle. Statements were rated as representing a low-high salience pattern, or a low salience topic and high salience vehicle; a similar salience pattern, with topic and vehicle demonstrating approximately equal levels of salience of the ground; or a high-low salience pattern, with a higher salience topic than vehicle.

Forty of the forty-five metaphorical statements were classified in the low-high or similar salience patterns. Eighteen of these were classified as low-high, while twenty-two were rated as similar. Again, the use of the sun as topic in many of the comparisons may have made low-high patterns more difficult for some students; 13 of the 22 similar salience statements came from the sun comparisons. These included such statements as “the sun is like a light bulb” and “the sun is like a burning stove.” Other similar salience statements, beyond the sun comparisons, included “a ball is as round as the moon,” “my cat is as soft as a pillow,” and “a tree is like a streetlight,” as well as a comparison of snow to sugar and a baby to a teddy bear.

Among the low-high salience patterns were many statements previously mentioned in other aspects of the analysis, such as the comparison of the sun to a ball of string used for weaving; the comparison of a classroom to a rainbow; and the comparison
of the rising and setting of the sun and moon to waking and sleeping patterns. Another clear low-high pattern appears in the following statement: “grass covers the mount/fire blazing on the ground.” This unfinished haiku demonstrates that the grass in question is dry and yellowed by the comparison to the high salience vehicle fire.

Three of the statements were rated as representing a high-low salience pattern, with the topic higher in salience than the vehicle on the specific ground. One of these was the flawed comparison of a tree to a pillow connected to a stick; the requirement to actively create the vehicle image in the mind in order to recognize the comparison resulted in the high-low rating. A second high-low pattern appeared in the statement that compared the sun to water based on the characteristic of being sometimes “nice and bright” and sometimes “hot and mean.” In this case, the characteristics of brightness and heat are more salient in the sun than in the vehicle water. The third comparison was close to an even rating, but is illustrative because of the need to add an adjective to the vehicle to bring the salience closer: “my blood is as red as a red crayon.” This statement utilizes a very high salience topic – indeed, an item that is often chosen as a vehicle based on its high salience of the characteristic “red” – and adds nothing to the understanding of the topic by comparing it to a vehicle that requires an adjective to possess the ground. Consequently, this statement was rated as high-low.

The final two statements were rated as unclassifiable. Again, these were the two vague metaphors in the tree comparison activity rated unclassifiable in other sections of the analysis as well. In this portion of the analysis, both had to be rated unclassifiable because even though the tree-monster comparison was clearly a visual attributional one, the specific ground to be emphasized was unclear.
In the following section, the notion of vehicle salience is explored further through analysis of another set of student products.

Selection of Vehicles

In the previous section, one of the activities included in the analysis was an activity in simile development in which students were given a ground and asked to create a simile around it. The activity as originally designed only required students to provide a vehicle for each ground, not a topic; the statements included above represent one teacher’s extension of the original activity. The intent of the original activity was to determine to what degree students could identify vehicles high in salience around the given ground, with reference to the need for high salience vehicles in effective metaphors as discussed above (Broderick, 1991; Glucksberg & Keysar, 1993; Ortony et al., 1984). Again, as noted in the previous section, this issue of salience represents a complex level of metaphor development, so the expectation for students in the study was for a fairly even mixture of high, medium, and low salience responses.

Student products from this vehicle development activity were collected from three of the five treatment classes (from Classes 1, 2, and 5). Student responses to the activity were tabulated and then classified in several ways. First, those responses that represented inappropriate vehicles for the given ground, or items that did not sufficiently bear the characteristic in question as to be useful in metaphors, were identified. The remaining responses were classified as high, medium, or low in salience with regard to the given ground. Responses were classified as high salience if the characteristic in question was a very prominent characteristic of the object; as medium if the characteristic might be likely to appear as a descriptor but would not be the most prominent or the best word to
describe the characteristic; and as low if the characteristic was an unlikely but not necessarily incorrect descriptor. A few additional details were noted about some responses: those responses that represented common vehicles for the given ground were identified (e.g., as quiet as a mouse), and those that required special knowledge of a book or a celebrity were also noted. Finally, a few responses that represented figurative comparisons rather than literal ones were identified. Specific student responses and their ratings are given in Appendix I; Table 23 summarizes results for each characteristic given in the activity.

Table 23

Summary of Selection of Vehicle Responses

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Number of Responses</th>
<th>High Salience</th>
<th>Medium Salience</th>
<th>Low Salience</th>
</tr>
</thead>
<tbody>
<tr>
<td>as tall as a ____</td>
<td>26</td>
<td>17</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>as strong as a ____</td>
<td>25</td>
<td>14</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>as quiet as a ____</td>
<td>27</td>
<td>21</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>as soft as a ____</td>
<td>27</td>
<td>26</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>as hard as a ____</td>
<td>27</td>
<td>23</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>as small as a ____</td>
<td>26</td>
<td>22</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>as happy as a ____</td>
<td>25</td>
<td>13</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>as mad as a ____</td>
<td>22</td>
<td>11</td>
<td>8</td>
<td>3</td>
</tr>
</tbody>
</table>
As Table 23 demonstrates, the majority of responses to each prompt were high in salience. The demand of the task, to relate an object to a characteristic instead of the more challenging task of relating two objects on the basis of a characteristic, perhaps suggests why students were able to provide high salience responses. In some instances, expressions in common use emerged from the task; these were rated as high salience responses even though students may have responded based on familiarity with the expressions rather than based on thinking carefully through the task. For example, for the prompt “as quiet as a ___,” 11 of the 21 high salience responses were “mouse”; in another example, 13 of the 23 high salience responses for “as hard as a ___” were “rock.” Other high salience responses offered somewhat more originality: “as small as a molecule,” “as beautiful as a rainbow,” “as quiet as a snake.”

As illustrated in Table 23, the more difficult items for students to find high salience responses were the two reflecting emotion (“as happy as a ___” and “as mad as a ___”) and the one asking for item reflecting the characteristic “strong.” With regard to “strong,” the prevalence of medium ratings is in part a reflection of student use of vehicles that are indeed strong but only in their resistance to being broken, a characteristic that is only sometimes reflected in the adjective “strong.” For example, several students used “iron” or “metal” as responses to the “strong” prompt. Although these items are certainly strong in some sense, the specific adjective is not necessarily the one that primarily comes to mind.
Similarly, in the case of the prompts reflecting emotion, many students provided vehicles that could be happy or mad, but these were not necessarily essential or primary characteristics. For example, for the “happy” prompt, several students offered the vehicle of “bird.” Although this may have been reflecting the idea that birds sing and therefore sound happy, the salience was rated as medium because happiness is not a characteristic of birds that comes immediately to mind when reflecting on them. Similarly, the use of “teachers” as a vehicle illustrating the characteristic “mad” may be an unfortunately common occurrence in some students’ experience, but in general “mad” is not one of the most salient characteristics that comes to mind about teachers.

Both of these discussions of medium ratings for student responses illustrate an important point about student performance on this task: as several researchers of young children and metaphor development have noted, sometimes the linguistic demands of metaphor tasks on students are more challenging than the metaphorical demands and may mask performance (Johnson & Pascual-Leone, 1989; Vosniadou, 1987). In several cases, responses that received a medium rating reflected the key characteristic to some degree, but had a slightly different adjective been used, the responses would have received a high salience rating. For example, in the “as tall as a ____” prompt, the response of “sky” received a medium rating, because generally we do not use the term “tall” to describe the sky. However, had the adjective in question been “high,” then “sky” would have been a high salience response and some of the other responses (e.g., “giraffe”) might have been rated as medium salience. This issue of matching appropriate adjectives to relevant vehicles suggests an important dimension for examination with regard to children’s production of metaphors.
Only a limited number of low salience responses appeared for each item, with no more than three rated as low salience in any category. In general, low salience ratings were given when the characteristic could apply to the response but only to a limited degree, in unusual cases, or when many other adjectives describing the object had been exhausted. For example, one student gave the response “as small as people.” Although people are certainly small in comparison to many objects and creatures of the world, it is rare that “small” would appear as a prominent characteristic of people in general. Another low salience response was “as hard as a volcano.” Again, although the sides of a volcano are generally hard, many other characteristics arise in the mind before hardness in thinking of a volcano. A third example was “as soft as Jupiter,” which was only given a rating at all because of the researcher’s familiarity with a popular children’s book in which characters land on Jupiter and sink into the “ground” to demonstrate to readers that the surface of Jupiter is not rock. However, clearly softness is not a prominent feature of Jupiter.

Related to this example of contextual familiarity defining a rating, another group of responses appeared which were not rated because the vehicle provided depended to such a degree on specialized knowledge that the degree of salience would be unknown to many readers. For example, for the prompt “as strong as a ___,” several students provided the names of specific wrestlers without clarifying that they were wrestlers, thus leaving it to the reader’s knowledge of celebrity wrestlers to understand the comparisons. One student used characters from a Roald Dahl book in response to nearly every question: “as quiet as the BFG,” “as soft as Sophie,” “as beautiful as Sophie,” “as bright as Sophie’s brain.” Although these examples represented good connections on the part of
the students to their own knowledge base, the comparisons could not be fairly rated in
terms of their salience because of the likelihood of limited familiarity on the part of many
readers. In a few cases, references to characters or celebrities were included in the rating
because of high visibility and familiarity of the names in the American culture; an
example of this is the comparison “as strong as Superman,” which received a high
salience rating. Some degree of researcher judgment naturally was necessary to determine
whether the vehicle provided was familiar enough to be rated or not.

In a few cases, student responses to the vehicle selection task were themselves
figurative, representing a more abstract level of attribution of adjectives. For example,
one student gave the response “as mad as a thunderstorm,” thereby ascribing emotional
characteristics to a natural phenomenon but nevertheless providing a high salience
vehicle. Two responses offered opposite sides of the same figurative idea: “as happy as a
sunny day” and “as bright as a happy face,” both of which require a level of figurative
understanding for use as high salience comparisons but are nevertheless strong images.
The following response created an unusual but strong image as well: “as quiet as time
freezing.”

These last few responses represent a figurative response to what was essentially a
literal task but clearly demonstrate these students’ capacity to utilize figurative and literal
meanings of words to express ideas. This capacity, along with the ability to recognize
speaker intent and the relationship between literal and figurative meanings of words, is an
important aspect of the development of figurative competence as noted in Levorato’s
(1993) model. In the following section, these abilities are explored further as they relate
to students’ responses to a metaphor interpretation task.
Metaphor Interpretation

The majority of student products collected in the study were the results of assignments for students to develop metaphorical comparisons themselves, while the two instruments used in quantitative analyses emphasized metaphor comprehension and interpretation. However, a closer qualitative analysis of student responses to metaphor interpretation tasks can also provide a clearer picture of the developmental capabilities of students in the sample.

Johnson and Pascual-Leone (1989) posited and tested a model of metaphor interpretation that classifies interpretations based on the type of mapping that occurs from vehicle to topic. According to this model, children progress developmentally through levels of processing characterized by the degree to which they are able to demonstrate how a ground existing at a high level of salience in the vehicle specifically applies to the topic. At early stages, children’s interpretation of metaphor often applies the ground in exactly the same way to topic and vehicle (Identity); they then progress to the ability to demonstrate an understanding of two slightly different meanings of the ground in the topic and vehicle (Analogy). Beyond this level, children reach a stage of interpretation at which they are able to elaborate more fully on the ground as it refers to the topic, either through offering a topic-specific example or instance of the ground (Concrete-Experiential Predicate) or a generic concept relevant to the topic (Generic-Conceptual Predicate). The model also incorporates a category of Inappropriate interpretations for those interpretations that do not correctly map the ground to topic and vehicle at all.

The authors noted that not all metaphors necessarily require the more advanced levels of interpretation; for metaphors based on immediate physical resemblance, for
example, Identity-level processing is sufficient. However, when more advanced level processing was required, they found that children's ability to produce more advanced interpretations as appropriate depended on age; their results demonstrated that although 7- and 8-year olds could generally handle Identity-level processing, Analogy-level did not frequently appear before age 9 or 10, while the Predicate levels were mostly found in children of at least age 11 or 12.

In this study, the products collected that most specifically required students to offer an interpretation of metaphor were the responses to question 2 of the LAIT pretest and posttest. On both forms of the test, this item required students to offer an explanation of a quote from a poem, where the quote was specifically a metaphorical statement. Quantitative results related to student scores on the item are discussed elsewhere in this chapter; in this section, the specific responses are examined more closely for classification within Johnson and Pascual-Leone's (1989) scheme. Since the sample represented a portion of the second grade population advanced in their verbal skills, it was predicted that the students would demonstrate some interpretations above the Identity level, potentially even demonstrating Predicate-level processing.

The metaphor students were asked to interpret on the pretest was a line from Carl Sandburg's poem "Fog," with the specific question given as follows: Use your own words to explain what you think the author means by the words "The fog comes in on little cat feet." For the posttest, students interpreted a line from Langston Hughes's poem "April Rain Song": Use your own words to explain what you think the author means by the words "Let the rain sing you a lullaby." Student responses to each question were reviewed and categorized first by similarity, because many offered similar interpretations,
and then rated according to the levels described above. Classifications were done separately for the pretest and posttest according to the specific responses related to each.

Table 24 demonstrates results by rating for the pretest and posttest; discussion specific to each and comparing the two follows.

Table 24

Classifications of Metaphor Interpretations

<table>
<thead>
<tr>
<th>Classification</th>
<th>Pretest</th>
<th></th>
<th>Posttest</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Inappropriate</td>
<td>42</td>
<td>50.6</td>
<td>24</td>
<td>30.8</td>
</tr>
<tr>
<td>Identity</td>
<td>38</td>
<td>45.8</td>
<td>16</td>
<td>20.5</td>
</tr>
<tr>
<td>Analogy</td>
<td>3</td>
<td>3.6</td>
<td>26</td>
<td>33.3</td>
</tr>
<tr>
<td>Concrete-Experiential</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>12.8</td>
</tr>
<tr>
<td>Predicate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generic-Conceptual</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2.6</td>
</tr>
<tr>
<td>Predicate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>83</td>
<td>100</td>
<td>78</td>
<td>100</td>
</tr>
</tbody>
</table>

Note. The numbers in the table represent combined numbers for the treatment and comparison groups. They are combined because this analysis is not intended to explore intervention effects but rather developmental performance, and because quantitative analysis yielded no significant differences between the groups in terms of performance on this test or this item.
The classification of Inappropriate was given to any response in this sample that offered a literal, incorrect interpretation of the metaphor. For example, a response to the pretest that indicated that fog actually has feet or one on the posttest that suggested that rain actually sings received an Inappropriate rating.

Identity responses for the pretest poem included any responses that employed descriptors that could apply equally to fog and a cat's feet with no clear distinction of differential attribution or primary focus on the fog. Examples of such responses included "softly," "slowly and quietly," "silently," and "gray like a cat's feet." These responses, although appropriate interpretations of the metaphor, do not clarify specific behaviors of the fog. As Table 24 indicates, nearly all of the pretest responses that were not rated Inappropriate were classified into the Identity category. Again, as Johnson and Pascual-Leone (1989) noted, not all metaphors require advanced level processing, which may be a factor in the limited number of responses beyond Identity level for this metaphor.

However, several students did produce Analogy-level responses, demonstrating an understanding of the ground of the metaphor as it specifically applied to the topic instead of applying equally to the topic and vehicle. These three responses all focused on the ground of "low" as a descriptor of fog: "the fog is low" (from two students) and "fog stays near the ground." These responses, through the choice of the terms "low" and "near," reflect a nuance of understanding about how fog is similar to but also different from a cat's feet, being low and near the ground but not on the ground as a cat's feet would likely be.

In the case of the posttest, the responses covered a broader range of the classifications. At the Identity level, posttest responses included those that simply
explained that “the rain sounds like a lullaby,” “the rain is soft like a lullaby,” or “the rain can make you feel sleepy.” When students referred to the actual falling of the rain more specifically, they were classified at the Analogy level: “when it falls on the roof it sounds like a lullaby,” “when the rain hits the ground it makes a sweet sound,” and “the sound rain makes when it splashes in puddles.”

Several student responses to the posttest question were classified at the Concrete-Experiential and Generic-Conceptual Predicate levels. At the Concrete-Experiential Predicate level, several of the students elaborated on their interpretation to describe the specific sound of the rain to which the metaphor referred. This provided a level of description that brought the interpretation beyond the Analogy level. Examples of this type of response included “the soft pat-pat rhythm was a lullaby,” “a steady beat pitter-patter that puts you to sleep,” “let the rain pitter patter you to sleep,” and “the rain makes music by going drip drop drip drop.” These responses demonstrate not only a clear understanding of how the ground applies somewhat differently to the topic and vehicle but also an ability to elaborate on how the ground applies to the topic specifically.

Two responses were classified at the Generic-Conceptual Predicate level because of their reference to broader concepts that can be applied to understanding of the rain. Both of these responses recognized that the vehicle of “lullaby” emphasized not only a certain sound of rain but also a type of rain that is not threatening but comforting. One of these responses was “when the rain comes let it sing to you or let the rain be comfortable.” Although the first part of this response actually reflects the criterion for classification as Inappropriate, the elaboration in the latter half of the statement supplies the conceptual understanding required for classification at this more advanced level. The
second Generic-Conceptual Predicate response was "pretend the rain is a lullaby and don't be afraid." This response goes beyond Identity-level processing with the comment to "pretend" the rain is a lullaby and then alludes to the conceptual dimension of the purpose of a lullaby and the characteristics of certain types of rain with the comment "don't be afraid."

Although the frequencies of responses within different classifications for the pretest and posttest were quite different, conclusions or generalizations related to the intervention or to maturity across the time span of the intervention for both groups would be premature. The two metaphors utilized in this analysis require different levels of processing for clear interpretation; Identity-level is more appropriate and likely an adequate level for understanding the pretest metaphor, while the Analogy level is somewhat easier to achieve in interpreting the posttest metaphor. However, the analysis does demonstrate that when presented with these two metaphors, many students were able to interpret them at a level consistent with or above predictions for their performance based on their age.

**Summary of Findings Related to Research Question 2**

The qualitative analysis portion of the study was intended to explore demonstrations within the sample of research-based patterns related to young children's production and interpretation of metaphor. The analyses of student metaphors, production of high-salience vehicles, and interpretation of metaphors in context demonstrated consistency with predicted patterns in most cases; in some instances, students demonstrated levels of performance found to be characteristics of older students in other studies, reflecting performance predicted for students two to three years older. Because
the students in the sample represent advanced verbal abilities relative to age, these performances exceeding age-related expectations were actually to be expected in this case. The variance of response across student products also demonstrates that the abilities of processing figurative expressions as units and using figurative language appropriately do indeed represent the more advanced levels of the Model of Figurative Competence and provide challenging experiences for young students.

*Findings Related to Intervention Effects*

Before addressing the two research questions that more specifically reflect attention to intervention effects, rather than developmental appropriateness more generally, a summary of findings related to the context of the implementation of the intervention is in order. The following pages outline details about the implementation gathered from classroom observations and teacher self-reports; the discussion will then turn to the findings related to Research Questions 3 and 4.

*Teacher Background and Unit Implementation*

In order to assess the teacher variable in relation to change in student performance, several data sources were used to secure information about the study teachers and their implementation of the intervention. Classroom observations were conducted in all treatment classes, and following unit implementation, teachers were asked to respond to a brief email questionnaire related to their background and implementation of the intervention. The specific questions asked of teachers are attached in Appendix J.
Teacher background.

The five teachers of the treatment group classes were all experienced teachers, with a range of 7-25 years of experience teaching. Most also had extensive experience teaching second grade, with a range of 5-23 years among four of the teachers. One teacher was teaching second grade for the first time; this teacher happened to be the one with the least overall teaching experience as well. One teacher had a master’s degree in gifted and talented education, while all had attended conferences and/or district workshops related to gifted and talented education. Each teacher taught in a different school in the district.

All five teachers had participated in training workshops on the language arts curriculum framework upon which the intervention unit was based, and all five had also attended the specific workshop related to the unit intervention used in this study. Before the study began, the teachers were given a list of which lessons out of the full unit they were required to teach for study participation and which products they should collect and submit to the researcher.

Teacher self-report of unit implementation.

Treatment group teachers were asked to report which of the unit lessons they taught in the time between pretest and posttest administration. Table 25 demonstrates the number of lessons reportedly implemented in each of the five treatment classes, with the percentage of the total required lessons thus represented.
Table 25

*Number and Percentage of Required Lessons Implemented by Treatment Group Teachers*

<table>
<thead>
<tr>
<th>Class</th>
<th>Number of Lessons Implemented</th>
<th>Percentage of Total Required Lessons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>11</td>
<td>100</td>
</tr>
<tr>
<td>Class 2</td>
<td>9</td>
<td>81.8</td>
</tr>
<tr>
<td>Class 3</td>
<td>9</td>
<td>81.8</td>
</tr>
<tr>
<td>Class 4</td>
<td>4</td>
<td>36.4</td>
</tr>
<tr>
<td>Class 5</td>
<td>10</td>
<td>90.9</td>
</tr>
</tbody>
</table>

*Note.* Class numbers correspond to school numbers given in Table 3 in chapter III.

As this table illustrates, most of the teachers implemented a high percentage of the required lessons, with four exceeding 80% and one teacher implementing all required lessons. One teacher, however, the teacher for Class 4, implemented only four of the required lessons within the time frame of the study. These results are all based on teacher self-report, and the question did not ask for specific detail on whether the lessons were fully implemented.

In addition to this information on lessons taught, teachers were also asked to discuss briefly their use of the two key tools included in the unit to encourage metaphor development, the Literature Web and the Metaphor Chart. All of the teachers indicated that they had used the Literature Web as specified within the unit. Two of the teachers indicated that they had used the web primarily as a group activity, with guidance and additional questioning from the teachers to support understanding. One teacher indicated that he saw a difference between the performance of his higher and lower ability students.
with the web, noticing more elaboration and discussion from high ability students: "My more able students were quite effective with it and the lower students had some difficulty with it and did not elaborate as much." This teacher noted that he had introduced the web early in the year for use outside of the specific intervention and that students had grown familiar with it through frequent use. Another teacher also explained that she had introduced the web prior to using the unit, and that she was able to see growth in student understanding of how to use it from the time it was introduced through the conclusion of the unit: "I think it is an excellent tool to analyze and interpret literature. When first introduced, the web was difficult for them to use. I don’t think they were used to really thinking about what they were reading. The web gave them a tool to use, so they could think about the reading in a more analytical way." These four teachers implied or stated in responses that the web was difficult at first for their students but became easier with repeated use at least for some students; the fifth teacher indicated that her students used the web with relative ease nearly from the beginning: "I thought that the kids did well with it, and had a clear understanding of how to use it. I do not think that it took them long to catch on."

With regard to the Metaphor Chart, four of the teachers indicated that they had used it as specified within the unit. Two mentioned making frequent references to the chart, even when not specified in the written unit, and one of these also indicated that students soon internalized the process of using the chart and would refer to it on their own when they recognized comparisons in text: "Once my students worked on the literature web a few times, they would refer to the [metaphor] chart when a comparison could be made. They internalized the concept of similar characteristics from our use of
the chart." One teacher indicated that he did not really use the chart while teaching the unit; this was the teacher who implemented only a few of the unit lessons as indicated in Table 25.

Teachers were also asked to indicate what other instruction in figurative language students had had during their second grade year prior to the implementation of the unit. Two teachers indicated that no emphasis on figurative language had occurred in previous instruction, while a third indicated that figurative language had been addressed when it arose in reading selections but not specifically emphasized. The remaining two teachers had taught some poetry earlier in the year; one had addressed poetry with emphasis on identifying main ideas and feelings, while the other had worked with story poems and locating similes within these poems.

Finally, teachers were asked to offer any additional comments on the unit or the implementation experience. The comments provided were generally positive, noting that the teachers found the activities of the intervention to be challenging and engaging for their students. Two teachers’ responses were particularly interesting to note. One of the teachers expressed a general concern that metaphor study is too complex for second graders, even advanced second graders, advancing the argument that, developmentally, these students are not capable of grasping the concepts involved: “I think that second graders are too young and concrete to really appreciate and use expressive language as freely as supposed. They are intelligent enough to understand the point during the lesson/reading, but not mature enough to apply it.” On the other hand, another teacher, who had only agreed reluctantly to participate, offered very positive comments at the conclusion of the intervention based on the progress and engagement she observed in her
own students: “The students responded with high level answers… They stimulated and challenged each other… I was very excited with their responses.” This teacher had also requested additional support during the implementation of the intervention through extra classroom visits by the researcher and, on one occasion, team teaching with the researcher. These differing teacher perspectives offered at the conclusion of the intervention offer interesting points for exploration with regard to student results.

Classroom observations.

Classroom observations in the study were conducted for two primary purposes: to ensure fidelity of implementation of the intervention, and to collect anecdotal data on student performance related to the metaphor comprehension and production tasks. The study design called for two observations to be conducted in each of the five treatment classes during the implementation period, for a total of ten observations. In actuality, only eight observations were conducted, with three teachers observed twice each and two teachers observed once each. Of the two teachers observed only once, second observations were scheduled for both. However, one of these teachers administered the posttest on the second observation date and engaged in no instruction from the unit. In the case of the other teacher, the researcher arrived for the observation and was informed on arrival that no instruction related to the intervention would occur that day. Timing, schedules, and distance prevented an additional observation from being scheduled.

All eight observations were conducted by the researcher, using the classroom observation form attached in Appendix F. This form is designed for use across the instructional unit and includes items related to all the major goals of the unit; because of this, not all items on the form are necessarily expectations of every lesson in the unit.
Consequently, teachers were assessed only on those items considered to be applicable to the particular lesson being taught.

"Meeting Expectations" was defined during the piloting process of the observation form as following the written lesson plan in terms of activities provided and questions asked, as well as using additional guiding questions to support student engagement within various teaching models. A rating of "Developing" related to the absence of intended activities or to the use of outlined teaching models ineffectively or inappropriately. A rating of "Exceeding Expectations" related to outstanding skill using the various teaching models, the incorporation of a number of additional salient questions, or the use of other teacher-selected materials or activities that clearly enhanced the written lesson.

Table 26 demonstrates the results of the classroom observations for the five teachers. Expected items were rated as Developing, Meeting Expectations, or Exceeding Expectations. One point was awarded for items Meeting Expectations and two points for Exceeding. No points were given for items rated as Developing.
Table 26

*Classroom Observation Scores for Treatment Group Teachers*

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Lesson Observed</th>
<th>Behaviors Expected</th>
<th>Point Score</th>
<th>Percentage Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>125%</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>7</td>
<td>6</td>
<td>86%</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>10</td>
<td>10</td>
<td>100%</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>10</td>
<td>9</td>
<td>90%</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>8</td>
<td>2</td>
<td>25%</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>6</td>
<td>5</td>
<td>83%</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>100%</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>9</td>
<td>7</td>
<td>78%</td>
</tr>
</tbody>
</table>

Total 66 57 86%

*Note.* Teacher numbers correspond to school numbers outlined in Table 3 in Chapter III.

Across all eight classroom observations, the majority of expected items were observed at a level of meeting expectations. A total of 66 behaviors were expected during the lessons observed, and total points across teachers added to 57, for a cross-observations total percentage score of 86%. The total score across classes was clearly affected by the low score of one of the teachers observed only once, who received points for only two of eight expected behaviors (25%). This low score reflects a poor implementation of the Literature Web with discussion that did not encourage students’ higher-level thinking. However, the teacher’s entire class period was not observed, and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
since it was the single observation conducted, generalizing about the teacher's overall performance related to the intervention is not really possible.

The scores for the three teachers observed twice all indicate a lower score for the second observation than the first. This finding is somewhat disturbing, given that one might reasonably expect to see performance gains as teachers gain more practice in using the intervention. In all three cases, the lower score on the second observation was related to the same item on the form, which was an item related to the emphasis placed on the generalizations about the unit concept. Although it is unfortunate that this behavior was lacking in the second observation, the specific item is less critical to the variables emphasized in the study than some of the other items on the form. On the other hand, in the case in which a teacher received a score exceeding expectations, the outstanding behavior related more specifically to the use of expressive language in a literature selection.

Summary of Findings Related to Teacher Background and Classroom Observations

The teachers in the treatment group were all experienced in teaching, and most were experienced in teaching second grade specifically. They all had also engaged in some professional development around gifted and talented education generally and the intervention specifically. The findings from the teacher questionnaire indicated consistency among four of the teachers around their implementation of the unit and its specific models as expected, although a fifth teacher completed only minimal implementation of the unit. In most cases, classroom observations also demonstrated consistency between expectations and implementation, although one teacher's implementation of the observed lesson was poor.
Despite this general consistency, however, the small number of teachers and the consequent potential for large variability warrant closer examinations of results at the classroom level. Moreover, the limited number of lessons taught by teacher 4 and the low observation rating received by teacher 3 suggested that some sub-analysis of student results excluding these teachers' classes would be justified. These sub-analyses appear within the explanation of findings in the following pages.

Findings Related to Research Question 3

Research Question 3 focused on exploring student learning gains on tests of metaphor comprehension based on exposure to the planned intervention. For each of the two instruments used as pre-post measures, analysis of covariance was computed to compare performance in the treatment and comparison groups. Findings related to each are outlined below.

Metaphor Comprehension Test results.

Three of the five treatment teachers and both comparison teachers returned sets of pre- and posttests of the Metaphor Comprehension Test (MCT). One treatment teacher misplaced her class's pretests during the course of the study and was unable to locate them, and another treatment teacher did not administer the pretests until after lesson 14 of the unit, so her results were excluded. The MCT was analyzed using analysis of covariance (ANCOVA). Levene's Test of Equality of Error Variances was run, with non-significant results supporting appropriate use of the procedure, although with the recognition that the large difference in sample size violates another assumption of the test. The results from the comparison between the treatment and comparison groups on the MCT are given in Table 27.
Table 27

*Analysis of Covariance for Treatment and Comparison Group on MCT*

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Pretest Mean (SD)</th>
<th>Posttest Mean (SD)</th>
<th>Adjusted Mean (SE)</th>
<th>F</th>
<th>Effect Size (η²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>35</td>
<td>15.56 (3.95)</td>
<td>14.54 (4.16)</td>
<td>14.05 (.58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>21</td>
<td>13.42 (4.49)</td>
<td>13.18 (4.31)</td>
<td>14.01 (.75)</td>
<td>17.00*</td>
<td>.39</td>
</tr>
</tbody>
</table>

* p < .001

The analysis of covariance yielded a statistically significant difference on the posttest score, using the pretest score as a covariate, with the treatment group scoring higher. Both groups actually demonstrated a drop in score from pretest to posttest, with a larger drop appearing in the treatment group, but the posttest mean for the treatment group was nevertheless significantly higher than that of the comparison group. The eta squared statistic indicated that the effect size was large (.39), indicating both statistical and educational significance, but the difference in sample size and the small size of the comparison group render any conclusions tentative.

In an effort to explore these results further, additional analyses were conducted on subsets of the sample. First, the scores for the teacher who admittedly taught only four of the expected lessons and did not use the Metaphor Chart were removed from the analysis. The results of this abbreviated comparison are given in Table 28.
Table 28

*Analysis of Covariance of Treatment and Comparison Groups on MCT, One Class Removed*

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Pretest Mean (SD)</th>
<th>Posttest Mean (SD)</th>
<th>Adjusted Mean (SE)</th>
<th>F</th>
<th>Effect Size (η²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>21</td>
<td>16.80 (3.72)</td>
<td>16.86 (2.89)</td>
<td>15.97 (.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>21</td>
<td>13.42 (4.49)</td>
<td>13.18 (4.31)</td>
<td>14.06 (.68)</td>
<td>18.18*</td>
<td>.48</td>
</tr>
</tbody>
</table>

*p < .01

In this analysis, the observed post-test score for the treatment group was slightly higher than the pretest score, although the adjusted mean again demonstrated a drop in score. Nevertheless, again, the treatment group’s adjusted mean score on the posttest is significantly higher than that of the comparison group (p < .01).

The troubling drop in scores for the treatment group revealed when all scores were included, along with the very small gain in scores with one class removed, suggested possibilities of instability in the instrument warranting further investigation. As discussed in Chapter III, piloting of the instrument had yielded statistically significant results in the test of equivalent-forms reliability and a small possibility of ceiling effect, but results for test-retest reliability on the posttest form were more questionable. Despite these issues, the forms were used because of the issue of time and because the equivalent-forms results were reasonable. However, in order to explore the study data further, the
issues of ceiling effect and equivalent form reliability were explored with the study
sample as well.

Ceiling effect, as discussed in the section addressing findings related to Research
Question 1, appeared to be more of a possibility with the study group than it had been
with the treatment group, with more than ten percent of students achieving scores above
the level of 85\% of the total possible score. Thus, ceiling effect is one possible
explanation for the minimal difference and decrease between pretest and posttest scores.
However, an additional possibility is the notion of regression toward the mean; using a
brief test assessing mastery with very similar forms administered as pretest and posttest,
regression toward the mean is one possible danger to validity of the study (Gall, Borg, &
Gall, 1996). The smaller standard deviation on the posttest for the comparison group,
suggesting a possibly diminishing variance, is one indication of the possibility of this
phenomenon at work in the analysis.

An assessment of equivalent-forms reliability was also conducted on the MCT
results in the study sample. Comparing results of the pretest and posttest across the entire
sample yielded a small but statistically significant correlation coefficient of .511 ($p <
.001$). Using the comparison group alone, this analysis yielded an even higher coefficient
of .679 ($p < .01$), although the small size of the comparison group causes these results to
be viewed with caution. Within the treatment group alone, the coefficient is smaller but
still statistically significant; the analysis yielded a correlation coefficient of .415 ($p <
.05$). Thus, similarly to the pilot results, the two forms of the test demonstrated somewhat
weak but statistically significant reliability, but the ceiling effect indicators were much
stronger in the study group than in the pilot group and may have affected the results.
To investigate the instrumentation issues further, a sub-analysis was conducted to determine results for different types of items on the MCT. For each of the six categories of task demand, as outlined in Table 4 in Chapter III, a correlational analysis was conducted to determine to what degree student scores on types of items related from pretest to posttest. For item types of which multiple items appeared on the test, scores were added to reach a total score for the given item type. This analysis was conducted with the treatment group scores and with the total sample scores, in each case including only students who had scores by item on both forms of the test, with the results demonstrated in Table 29. The results indicate small but significant correlations for some item types, particularly within the larger entire sample group. However, more than half of the categories did not show significant correlations from pretest to posttest for the treatment group, and only three of six showed significant correlations for the entire sample, suggesting that pretest and posttest performance do not necessarily represent effective predictors of one another on the instrument.
Table 29

*Pretest – Posttest Correlations by Item Type – MCT*

<table>
<thead>
<tr>
<th>Item Type</th>
<th>Number</th>
<th>n</th>
<th>Pearson Correlation (Treatment Group only)</th>
<th>n</th>
<th>Pearson Correlation (Entire Sample)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifies topic of metaphor</td>
<td>2</td>
<td>34</td>
<td>.188</td>
<td>55</td>
<td>.206*</td>
</tr>
<tr>
<td>Identifies topic of given vehicle</td>
<td>2</td>
<td>31</td>
<td>-.028</td>
<td>52</td>
<td>.220</td>
</tr>
<tr>
<td>Identifies key characteristics of</td>
<td>3</td>
<td>32</td>
<td>.357*</td>
<td>50</td>
<td>.352*</td>
</tr>
<tr>
<td>topic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identifies vehicle of given topic</td>
<td>2</td>
<td>32</td>
<td>.317</td>
<td>52</td>
<td>.399**</td>
</tr>
<tr>
<td>Identifies key characteristics of</td>
<td>3</td>
<td>30</td>
<td>.492**</td>
<td>49</td>
<td>.511**</td>
</tr>
<tr>
<td>vehicle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identifies main idea of selection</td>
<td>1</td>
<td>31</td>
<td>.126</td>
<td>52</td>
<td>-.011</td>
</tr>
</tbody>
</table>

* *p < .05

** **p < .01

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Pre-test and post-test mean scores by item type were then computed for the treatment and comparison groups for descriptive comparison. These results are reported in Table 30.

Table 30

*Item Analysis – MCT*

<table>
<thead>
<tr>
<th>Item</th>
<th>Treatment Group</th>
<th>Comparison Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest</td>
<td>Posttest</td>
</tr>
<tr>
<td>Identifies topic of metaphor (possible score: 2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>1.23</td>
<td>1.51</td>
</tr>
<tr>
<td>$SD$</td>
<td>.62</td>
<td>.65</td>
</tr>
<tr>
<td>$n$</td>
<td>34</td>
<td>37</td>
</tr>
<tr>
<td>Identifies topic of given vehicle (possible score: 4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>2.75</td>
<td>2.35</td>
</tr>
<tr>
<td>$SD$</td>
<td>.92</td>
<td>1.47</td>
</tr>
<tr>
<td>$n$</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>Identifies key characteristics of topic (possible score: 6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>4.77</td>
<td>4.32</td>
</tr>
<tr>
<td>$SD$</td>
<td>1.40</td>
<td>1.66</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
The item analysis demonstrates that for most item types, the mean scores changed in a similar pattern from pretest to posttest for the treatment and comparison groups, with only the third item type ("Identifies key characteristics of vehicle") demonstrating an increase in one group and a decrease in the other. For three of the six item types, both the
treatment and comparison groups saw a decrease in scores from pretest to posttest, while a fourth item type resulted in a drop for the treatment group. These item-level results, both the correlational analyses and the descriptive comparisons, support the notion of an existing but weak reliability between the forms, as well as again illustrating possible sources of ceiling effect.

Further analyses were conducted to explore patterns within the treatment group. Subanalyses were conducted to determine to what degree results were related to demographic factors such as class group, age, and gender. A test assessing the correlation between age and gain scores on the instrument yielded a nonsignificant Pearson coefficient of -.332. In an analysis of gender differences on the test, both gender groups showed a drop in scores from pretest mean to posttest mean, but the drop in score for girls was much smaller than that for boys. Sample size for the gender comparison was too small for an analysis of covariance to be justified; descriptive statistics for the gender groups are reported in Table 31.

Table 31

*Descriptive Statistics for Gender on MCT (Treatment Group)*

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Pretest Mean (SD)</th>
<th>Posttest Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys</td>
<td>19</td>
<td>15.78 (3.49)</td>
<td>14.04 (4.45)</td>
</tr>
<tr>
<td>Girls</td>
<td>16</td>
<td>15.30 (4.54)</td>
<td>15.14 (3.83)</td>
</tr>
</tbody>
</table>

Because the sizes of individual class groups were so small and ranged so widely, tests of significance were not run to compare scores across class groups. Mean pretest
and posttest scores for each class group are given in Table 32, with posttests included for
the two classes not submitting pretests.

Table 32

Pretest and Posttest Means by Class Group – MCT

<table>
<thead>
<tr>
<th>Class</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Gain Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>n</td>
</tr>
<tr>
<td>1</td>
<td>17.08</td>
<td>4.33</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>16.42</td>
<td>2.94</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>15.24</td>
<td>4.10</td>
<td>23</td>
</tr>
<tr>
<td>4</td>
<td>13.70</td>
<td>3.65</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>12.47</td>
<td>3.16</td>
<td>9</td>
</tr>
</tbody>
</table>

The scores demonstrate considerable differences among the class groups in terms
of scores and gains or losses. Class 4, the class with the lowest posttest score and the
largest drop in scores, was the class removed from the analysis given in Table 28. Classes
1 and 2 clearly scored much higher than Class 4 on the pretest, with the gap between
them widening at the top and bottom. Moreover, the score changes from pretest to
posttest are similar for classes 1 and 2 but demonstrate a large drop for class 4. The issue
of ceiling effect, however, clearly emerges more prominently from Classes 1, 2, and 3
than from Classes 4 and 5. Again, this class-level analysis demonstrates that the high
mean score on the test generated by the scores of Classes 1 and 2 may have had an
important influence over the results of the analysis of covariance.
Literary Analysis and Interpretation Test results.

All five treatment teachers and both comparison teachers returned sets of pre- and post-tests for the Literary Analysis and Interpretation Test (LAIT). The results from the comparison between the treatment and comparison groups on this instrument are given in Table 33. Both groups demonstrated increases in scores from pretest to posttest, with a larger gain score for the treatment group, but the analysis of covariance revealed the difference between groups to be nonsignificant.

Table 33

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Pretest Mean (SD)</th>
<th>Posttest Mean (SD)</th>
<th>Adjusted Mean (SE)</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>55</td>
<td>9.27 (2.38)</td>
<td>11.35 (2.37)</td>
<td>11.36 (.34)</td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>18</td>
<td>10.67 (3.07)</td>
<td>10.78 (2.67)</td>
<td>10.75 (.59)</td>
<td>.387</td>
</tr>
</tbody>
</table>

As with the MCT, additional analyses were conducted to explore results further. A second ANCOVA was conducted with the class completing only a few lessons removed. However, results of this analysis were similar to those found with the entire treatment group, yielding a nonsignificant F value of .41 (p = .67). Although neither ANCOVA demonstrated significant differences between treatment and comparison groups, paired-sample t-tests were run on the treatment group and comparison group separately to compare pre- and post-test scores for growth gains, with the recognition that
results could not be attributed to the intervention because of the lack of significant differences between groups. The $t$-test for the treatment group yielded significant gains, while the test for the comparison group was nonsignificant, as displayed in Table 34.

Table 34

*Pretest-Posttest Differences on LAIT in Treatment and Comparison Groups*

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Treatment</td>
<td>9.27</td>
<td>2.38</td>
</tr>
<tr>
<td>Comparison</td>
<td>10.67</td>
<td>3.07</td>
</tr>
</tbody>
</table>

*$**p < .001$

These results are interesting in that they reinforce gain score differences apparent at a descriptive level between the treatment and comparison groups, although again no conclusions related to intervention effects are justified. Moreover, a correlational analysis between the two forms of the LAIT also yielded nonsignificant correlation coefficients, again demonstrating potential instrumentation issues. This result and the nonsignificant ANCOVA prevent conclusions regarding growth gains on the instrument, but the results are nevertheless interesting and suggestive of possibilities for further investigation.

Within the treatment group, again as with the MCT, comparisons were conducted to explore differences related to gender, age, and class group. A correlational analysis assessing the relationship between gain scores on the test and age yielded a nonsignificant Pearson coefficient of -.030. Analysis of covariance based on gender yielded a nonsignificant result, although again at a descriptive level females scored slightly higher than males on the posttest, as demonstrated in Table 35.
Table 35

Analysis of Covariance by Gender on LAIT (Treatment Group)

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Pretest Mean (SD)</th>
<th>Posttest Mean (SD)</th>
<th>Adjusted Mean (SE)</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys</td>
<td>25</td>
<td>9.44 (2.20)</td>
<td>10.96 (2.52)</td>
<td>10.93 (.47)</td>
<td></td>
</tr>
<tr>
<td>Girls</td>
<td>30</td>
<td>9.13 (2.56)</td>
<td>11.67 (2.23)</td>
<td>11.69 (.43)</td>
<td>1.72</td>
</tr>
</tbody>
</table>

Comparison across the class groups within the treatment sample demonstrated that all five groups showed gains from pretest to posttest, though the degree of gain was not consistent across groups. These results are shown in Table 36.

Table 36

Pretest and Posttest Means by Class Group - LAIT

<table>
<thead>
<tr>
<th>Class</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Gain Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>n</td>
</tr>
<tr>
<td>1</td>
<td>7.78</td>
<td>2.11</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>11.11</td>
<td>2.03</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>8.50</td>
<td>2.04</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>9.25</td>
<td>2.72</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>9.33</td>
<td>1.73</td>
<td>9</td>
</tr>
</tbody>
</table>
It is notable in this analysis by class that the descriptive results do not align so clearly with observational and implementation details of the five teachers as they did with the MCT. The two classes with teachers who demonstrated some weakness in observations or reported implementation (classes 3 and 4) demonstrated posttest and gain score results in the middle of the treatment group, not so clearly separated from the teachers with more complete implementation.

The LAIT was also sub-analyzed by item. This test consisted of four items, as outlined in Chapter III. Pretest and posttest means for each item in the treatment and comparison groups are given in Table 37.

Table 37

*Item Analysis – LAIT*

<table>
<thead>
<tr>
<th>Item</th>
<th>Treatment Group</th>
<th>Comparison Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest</td>
<td>Posttest</td>
</tr>
<tr>
<td><strong>Item 1 – main idea</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>M</em></td>
<td>2.03</td>
<td>2.33</td>
</tr>
<tr>
<td><em>SD</em></td>
<td>.67</td>
<td>.75</td>
</tr>
<tr>
<td><em>n</em></td>
<td>63</td>
<td>60</td>
</tr>
<tr>
<td><strong>Item 2 – analyze quote</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>M</em></td>
<td>2.92</td>
<td>4.03</td>
</tr>
<tr>
<td><em>SD</em></td>
<td>1.24</td>
<td>1.50</td>
</tr>
<tr>
<td><em>n</em></td>
<td>63</td>
<td>60</td>
</tr>
</tbody>
</table>
Item 3 — change in understanding of topic

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.94</td>
<td>.80</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>2.87</td>
<td>1.24</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>2.60</td>
<td>1.47</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>2.89</td>
<td>1.57</td>
<td>18</td>
</tr>
</tbody>
</table>

Item 4 — new title

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.19</td>
<td>1.00</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>2.13</td>
<td>.62</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>2.50</td>
<td>1.10</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>2.22</td>
<td>.65</td>
<td>18</td>
</tr>
</tbody>
</table>

The table demonstrates that the gain scores within the treatment group resulted primarily from questions 2 and 3. The comparison group also demonstrated positive gains on these items, though the gains were smaller than they were for the treatment group. Notably, items 2 and 3 are the test items most closely aligned with the intended instruction surrounding the structure and purpose of metaphor, relating to the Metaphor Chart and discussion questions around given literature selections. Furthermore, in order to achieve a score of 4 on question 2, the student must make reference to the ground of the metaphor in the response, which is unstated in the text of the question. Thus, the treatment group’s achievement of a mean score just above 4, even though it is not statistically higher than that of the comparison group, may represent some advancement in demonstration of metaphorical thinking.
Summary of Findings Related to Research Question 3

The data from both instruments administered on a pre-post basis provided some significant findings but also left some questions unanswered and raised additional ones. On the Metaphor Comprehension Test, although the treatment group outperformed the comparison group on the posttest at a statistically significant level, a decrease in the treatment group’s mean score from pretest to posttest raises questions related to the stability of the instrument. Further analysis of results on this basis demonstrated a possibility of ceiling effect on both forms of the instrument. The Literary Analysis and Interpretation Test yielded significant gain scores for the treatment group when examined alone, but the results were not significant in relation to the comparison group; thus, the gains cannot justifiably be attributed to the intervention. However, on the LAIT, item analyses at a descriptive level suggested a possible relationship between the emphasis of the instruction in the intervention and student gain scores within the treatment group.

Within the treatment group, large differences in performance existed among classes on the MCT and appeared to reflect differences in reported implementation among the classes; smaller differences among classes were evident on the LAIT and were less clearly connected to teacher differences. Additional demographic analyses within the treatment group yielded small gender differences that were significant for the MCT but not for the LAIT; age differences were not significant for either test.

Findings Related to Research Question 4

Research Question 4 focused on relating student scores on the written assessments utilized in the study to teacher assessment of performance as determined by observations.
of participation in the activities of the intervention. The Teacher Assessment of Student
Performance (TASP) included ten items based on the definitional structure of metaphor
and the abilities detailed in the conceptual framework. Each item was to be rated on a
scale of 1 to 4, and teachers could exclude items they felt unable to assess. This exclusion
option was exercised by teachers for a number of students on item 10, which concerned
student ability to use figurative language spontaneously. As this was the only item
excluded by teachers, and because it dealt with a level of figurative competence not
assessed by the two instruments, totals for the form were computed based only on items
1-9, for a total possible score of 36. The mean total score across the treatment group on
this form was 28.4, with a standard deviation of 7.91. Because the form was so dependent
on the interpretation of individual teachers as well as the performance of individual
students, means were computed for the five treatment classes separately and are reported
in Table 38.

Table 38

<table>
<thead>
<tr>
<th>Class</th>
<th>n</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>27.50</td>
<td>4.40</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>33.67</td>
<td>2.69</td>
</tr>
<tr>
<td>3</td>
<td>22</td>
<td>33.23</td>
<td>2.78</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>17.94</td>
<td>7.74</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>31.50</td>
<td>5.35</td>
</tr>
<tr>
<td>Total</td>
<td>67</td>
<td>28.4</td>
<td>7.91</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
This display of mean scores by class suggests that in several cases, teacher ratings of their students approached the ceiling for the form; indeed, 18 students (26.9%) were scored full points on the form, and over half of the students (35; 52.2%) were scored above 30 or above 85%. The degree to which teachers carefully observed each student with regard to the specific test items is questionable, particularly in the classes with high means and small standard deviations. Indeed, one teacher (Class 3) had to be asked repeatedly at the conclusion of the study to submit her forms, and it is likely that this teacher actually completed the forms more than a month after completing the unit. The teacher with the lowest mean score and largest standard deviation above (Class 4) was the teacher who had implemented only a few lessons; his responses to the form indicated a closer student-by-student examination of the items, but the degree to which the lessons he implemented had provided opportunity for observing the behaviors is questionable. Nevertheless, this comparative placement of Class 4’s mean as the lowest in the treatment group echoes its mean performance on the MCT posttest as the lowest scoring class and the LAIT posttest as the second lowest (see Tables 32 and 36), with the caveat that the class comparisons were conducted only at a descriptive level.

An additional interesting point to note about the TASP results is that the teacher who argued in her commentary on the intervention that it was developmentally too advanced for the students was the teacher of Class 5, which achieved the third highest mean score on the TASP, within five points of the ceiling. However, Class 5’s mean scores on both posttests were also at or near the bottom in the comparisons across the treatment classes. Thus, the exploration of relationships between the TASP items and
other results given below is done with caution and recognition of the subjectivity in teacher interpretation of the form and context.

TASP scores were correlated with student scores on both posttest measures to assess potentially important relationships. The correlation of the TASP scores to posttest scores on the MCT yielded a statistically significant correlation coefficient of .475 ($p < .001$). The relationship between the TASP scores and the LAIT scores, on the other hand, yielded a nonsignificant correlation coefficient of .166 ($p = .22$).

Because of the close design connection between items on the TASP and item type on the instruments, specific item correlations were also conducted. Three specific items on the TASP were correlated with groups of items on the MCT. Five of the six item types from the test were collapsed into three categories to relate to the specific items on the TASP. Scores for these three categories were then computed by adding student posttest scores for the items in question. Results indicated small but statistically significant correlations between teacher assessment of these abilities and student performance on the MCT, as displayed in Table 39.

Table 39

*Item Correlations Between TASP and Item Groups on MCT*

<table>
<thead>
<tr>
<th>TASP Item</th>
<th>MCT Item Groups</th>
<th>Items per</th>
<th>Correlation</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. ability to identify topic of a figurative expression</td>
<td>Topic – identify topic of metaphor, identify topic of given vehicle</td>
<td>4</td>
<td>.311*</td>
<td>63</td>
</tr>
<tr>
<td>Ability to Identify</td>
<td>Vehicle of a Figurative Expression</td>
<td>Ability to Identify</td>
<td>Ground of Comparison in a Figurative Expression</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------------</td>
<td>---------------------</td>
<td>-----------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>5. Ability to Identify</td>
<td>Vehicle – identify</td>
<td>2</td>
<td>0.249*</td>
<td>67</td>
</tr>
<tr>
<td>6. Ability to Identify</td>
<td>Ground – identify</td>
<td>6</td>
<td>0.499**</td>
<td>64</td>
</tr>
</tbody>
</table>

* $p < .05$

** $p < .01$

Two additional items on the TASP were analyzed with the final item type from the MCT and with specific items from the LAIT. The items from the TASP addressing student ability to infer author purpose and to process large amounts of text at once to analyze meaning were related to the item assessing main idea on the MCT and to individual items on the LAIT. None of these comparisons showed statistically significant correlations, as demonstrated in Table 40.
Table 40

*Correlation Coefficients for Relationships of TASP Items 7 and 8 to Test Items*

<table>
<thead>
<tr>
<th>Teacher Assessment Item</th>
<th>MCT</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>main idea</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>question</td>
<td>.092</td>
<td>.135</td>
<td>.038</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. ability to infer author</td>
<td>-.009</td>
<td>.160</td>
<td>.092</td>
<td></td>
<td></td>
</tr>
<tr>
<td>purpose regarding topic of a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>figurative expression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. ability to process large</td>
<td>-.046</td>
<td>.038</td>
<td>.185</td>
<td>.099</td>
<td>-.086</td>
</tr>
<tr>
<td>amounts of text at once to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>analyze meaning of a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>figurative expression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conclusions regarding these analyses of the relationship between specific items on the TASP and items or item groups on the two posttest measures are made with additional caution because of the potential for error with so few items and the limited variability on any given item across the TASP scores.

*Summary of Results Related to Research Question 4*

The TASP was employed in the study to gain another source of student performance beyond what could be illustrated by tests and a few products. Because of the heavy emphasis on class discussions and because it was impossible for the researcher to be present for a majority of class sessions, the TASP provided the opportunity for teachers to give additional details regarding student performance. Scores on the TASP...
were generally high, with mean scores approaching the ceiling in several classes; these results may indicate a lack of differentiation of response among the teachers. Because of the potential for subjectivity in teacher responses, results related to the TASP are treated with caution. Statistically significant correlations between the TASP and student posttest scores on the MCT appeared, both in terms of total scores and specific item relationships. Correlations between the TASP and the LAIT were not statistically significant either at a total score or item score level.

Summary of Findings

The research findings of this study of metaphor development fall into two major categories. The first of these relates to the development of figurative competence and demonstrations of that development as they emerged within the study, including the relationship between figurative competence and ability. The second relates to the quasi-experimental aspect of the study and the effects of the intervention on demonstrations of metaphor development.

Key Findings Related to the Development of Figurative Competence

1. Students in both the treatment and comparison groups demonstrated the ability to respond to tasks requiring comprehension, interpretation, and production of metaphor, as demonstrated through performance on the instruments utilized in the study that included a range of scores in some cases approaching the ceiling for the instruments. Teacher assessments of student performance also indicated developmental readiness, although one teacher's commentary suggested that the content and tasks were too complex for second grade students.
2. Within the limited range of ages encompassed by the sample, chronological age did not correlate significantly with performance on any study tasks. Females scored slightly higher than males on the study instruments, according to descriptive data, but statistically significant differences did not appear based on gender.

3. Performance on the MCT yielded a small but statistically significant positive correlation with verbal reasoning ability as measured by a subtest of the TCS, while a small but significant negative relationship appeared between gain scores on the LAIT and the verbal reasoning subtest, and an even smaller but significant and positive correlation emerged between posttest scores on the LAIT and the memory subtest. The potential for ceiling effect for students in the sample on the TCS was noted, however, based on high mean scores.

4. Student products demonstrated several predicted patterns related to the development of metaphor production, including a higher likelihood of flawed than vague metaphors, in addition to producing a number of appropriate metaphors; preference for using attributional over relational comparisons; frequent use of visual characteristics as the grounds for metaphors; and avoidance of low salience vehicles but no clear preference for asymmetry over symmetry in terms of salience. However, the product analysis did yield results suggesting that the study sample exceeded age expectations in the areas of including only a small number of nonmetaphorical or vague comparisons on metaphor tasks; several complex metaphors incorporating attributional and relational characteristics; a frequent use of abstract relationships as the grounds for metaphors; and a high percentage of high salience vehicles in metaphors.
5. Student products related to metaphor interpretation also yielded predicted patterns and some results exceeding expectations in terms of the level of semantic mapping used to explain the meaning of a metaphor. A high percentage of Inappropriate and Identity-level responses appeared, reflecting age expectations, but a number of Analogy-level responses and several Predicate-level responses exceeded age-related expectations for metaphor interpretation.

Key Findings Related to the Effects of Instruction

1. The treatment group outperformed the comparison group on the MCT posttest at a statistically significant level, although the treatment group alone demonstrated a drop in mean score on the instrument from pretest to posttest. The removal of one treatment class whose participation in the intervention had been limited left the treatment group with a small gain from pretest to posttest on the MCT and stronger statistical significance compared to the comparison group. On the LAIT, the treatment group alone demonstrated statistically significant gains, but the results were not statistically significant when compared to the comparison group’s scores.

2. Results of the MCT revealed a possible ceiling effect at work on both forms of the instrument across the treatment and comparison conditions; these results exceeded a weaker suggestion of ceiling on the pretest form only suggested by pilot results.

3. Item analyses for the two instruments at a descriptive level indicated similar scoring patterns for the treatment and comparison groups. However, on the LAIT, item scores indicated much larger gains on some items for the treatment group than the comparison group, particularly for the two items most closely aligned with the instructional tools of the intervention.
4. Classroom observation and teacher self-report data demonstrated consistent implementation of required lessons and key instructional tools across most treatment group teachers, although limited implementation by one teacher and a weak performance by another on a classroom observation were noted.

5. Descriptive statistics organized by class revealed large differences in performance within the treatment group on the MCT and somewhat smaller yet still noticeable differences within the group on the LAIT. In some but not all cases, these differences reflected differences reported by teachers regarding their implementation of lessons and their reactions to the intervention.

6. Data analysis of TASP results yielded high scores and limited differentiation of perceived student performance in most treatment classes, though results indicated a statistically significant correlation between the TASP and the MCT. The correlation between the TASP scores and LAIT scores was not statistically significant. Item-level analyses also yielded statistically significant correlations between MCT item groups and relevant questions on the TASP.

The following chapter discusses these findings further, drawing some conclusions based on the results and suggesting implications for future research and practice.
CHAPTER V
Discussion, Conclusions, and Implications

This study was organized around four major research questions, with two underlying dimensions of focus. One of these dimensions was the development of figurative competence in young children, explored with emphasis on the developmental readiness of verbally talented second grade students to interpret and produce metaphorical comparisons. Emerging from this first dimension was the second, larger dimension of the study, which explored the question of whether this developmental process can be influenced by instruction that provides specific scaffolding related to the structure of metaphor, in line with Vygotsky's (1986) notions of cognitive mediation. All of the study instruments were primarily intended to address one or both of these topics, as well as to provide data supporting the exploration of relationships between students' demographic and ability characteristics and their performance on the study instruments.

The results addressing these two central dimensions of the study, as demonstrated through the findings discussed in Chapter IV, will be discussed in the pages that follow, with the goal of demonstrating conclusions to be drawn from this research and implications for future research and practice. In addition, implications reflecting unforeseen findings emerging from the implementation process will be discussed as they relate to school-based research endeavors and classroom practice.

The Development of Figurative Competence

Studies of young children and metaphor development across the last 20 years have demonstrated that many children under the age of ten are capable of understanding
metaphors and offering interpretations of them, more or less effectively depending on the conditions of presentation and the specific task demands (Broderick, 1991; Johnson & Pascual-Leone, 1989; Vosniadou, 1987). With regard to metaphor production, the research is somewhat less consistent, because of the difficulty in many cases of assessing the variable of intent in children’s metaphors (Vosniadou, 1987; Winner, 1988). Nevertheless, indications exist that some children in the primary grades are capable of producing metaphors, although the ability to produce figurative language deliberately appears to develop later than the ability to comprehend and interpret figurative expressions (Broderick, 1991; Levorato, 1993; Levorato & Cacciari, 1995). This notion of a developmental progression from comprehension to production is supported in most of the studies of metaphor development, as demonstrated in Levorato’s Model of Figurative Competence (Levorato, 1993; Levorato & Cacciari, 1995), described in Chapters II and III.

The results of this study are consistent with earlier research in demonstrating that second grade students, between the ages of 7 and 9, can comprehend metaphors and offer interpretations of metaphorical statements, although the sophistication of responses they offer varies across individuals and across prompts. The two major instruments utilized in the study, the MCT and the LAIT, both required that students be able to interpret metaphorical statements within poetic contexts in order to respond to questions. Mean scores for both the treatment and comparison groups on the MCT pretest and posttest were above 50%, suggesting that the students were capable of understanding the metaphors presented in the poems and responding correctly to questions about them. Similarly, on the LAIT, mean posttest scores for both groups on the questions most
directly tied to metaphorical understanding were at a medium level in terms of the scoring rubric, also suggesting that the task demands were neither too easy nor too difficult for the range of students in the group. Moreover, the assessment of student products demonstrated that in terms of both comprehension and production, students in the study group performed at or above predicted developmental levels.

The task demands of the study instruments and the products collected required students to demonstrate several of the developmental abilities identified in the conceptual framework as representative of performance at approximately ages 7 to 9. These abilities included the ability to understand relationships between different meanings of words, including both literal and figurative meanings, and the ability to suspend a purely referential strategy. In addition, many of the task demands required that students process entire figurative expressions at once to provide correct interpretations of metaphors, which is an ability expected at a somewhat later stage of development according to the conceptual framework. The demonstration of this ability among some students in the sample, as explored specifically in the LAIT results, indicates that for students of advanced verbal skills, interpretation abilities may appear earlier than the developmental model predicts. Therefore, metaphor tasks that require students to supply interpretations instead of merely selecting interpretations from a set of possible responses may, in fact, represent appropriate task demands for this population.

*Developmental Appropriateness and Student Ability Differences*

This study was designed to focus on a specific group within the second grade population of the participating schools, including only those students demonstrating advanced reading skills for their grade level. However, within the population of high
ability students there nevertheless exists considerable variability in terms of verbal aptitude and achievement. Student performance on the study instruments demonstrated a range of performance, including some students who approached ceiling scores on the MCT in particular as well as students who demonstrated scores much lower. The qualitative analysis, as well, demonstrated in each of the aspects examined a range of performance from below expectations for the age level to a higher performance than might have been expected. The quantitative analysis conducted to assess the statistical significance of correlations between scores on study instruments and scores on the TCS was intended to explore this range of performance further.

The study hypotheses predicted that a statistically significant relationship would exist between scores on the study instruments and scores on other tests of verbal and/or general ability and achievement. This hypothesis was supported only to a very limited degree, with statistically significant correlations between the verbal reasoning subtest of the TCS and posttest scores on the MCT and between this same subtest and gain scores on the LAIT. The small correlations support the notion that the tests may be drawing upon similar or related skills, which is a reasonable assumption to be made about multiple tests requiring verbal reasoning. However, the small size of the correlation coefficients, the clustering of TCS mean scores near the ceiling for each subtest, the evidence of ceiling effect on the MCT, and particular features of each analysis raise interesting questions for consideration.

In the case of the MCT, the significant relationship is a positive one, though small, demonstrating a small degree of predictability of one set of scores based on the other. However, as demonstrated in Chapter IV and discussed in more detail in the pages

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
that follow, other analyses related to the MCT revealed potential instability in the instrument, particularly an issue of possible ceiling effect, which is also a possibility on the TCS within the study sample, or of regression effect. The drop in scores from pretest to posttest within the treatment group is a curious and troubling aspect of the findings, and consequently any conclusions related to gain scores or posttest scores on this instrument must be made with extreme caution. Nevertheless, because the relationship was statistically significant, the findings support the notion that metaphor comprehension in verbally talented second grade students, as measured by the MCT, may be tentatively predicted by performance on the TCS verbal reasoning subtest.

The findings related to the LAIT are also curious, in that gain scores but not posttest scores demonstrated a significant correlation with the verbal reasoning subtest of the TCS, and also in that the statistically significant relationship was a negative one. This result seems anomalous in some ways in its suggestion that the more verbally able the student, the less capable he or she may be of improving performance on a specific verbal instrument; however, another way of interpreting the same result is that less verbally able students had more room to improve and thus had a stronger likelihood of demonstrating gains. The result and these interpretations also raise questions about the adequacy of the instruments, however; since gain scores are the variable in question, the requirement that the forms of the test be equivalent is even more important, and the analysis of the two forms of the test with the study sample yielded a nonsignificant correlation coefficient. On the other hand, the possibility that the scores on the TCS were limited and compressed by the ceiling provides another possible explanation for the result, especially since the gain scores showed much more variability in terms of standard deviation than
the TCS scores. With regard to the significant correlation between posttest scores and the memory subtest of the TCS, there are also interesting possibilities for consideration, since the memory test demonstrated the lowest mean scores of the four TCS subtests, suggesting the possibility of a more limited influence of any ceiling effect. The positive correlation may support some relationship between ability as measured by the TCS and performance; however, without significant results from the tests more related to the same domain of performance as the LAIT and with such a small correlation, such conclusions would be difficult to justify.

Despite the limited results of the regression analyses, on a descriptive, anecdotal level, several of the teachers commented that the task demands of the intervention had been differentially challenging for students they perceived to be of different levels of ability in their classes. Although one teacher expressed her opinion that the study of metaphor is too advanced for second grade students, even those of advanced verbal abilities, the other four indicated that the metaphor tasks were challenging but not inappropriately so for their students. Nevertheless, despite the references to the different levels of challenge provided by the tasks for different students, three of the five teachers’ TASP results reached mean scores above 31 on the 36-point instrument, with two of these three showing relatively low standard deviations. Interestingly, the teacher who argued that the study of metaphor was too difficult was one of the three with a mean TASP score above 31, while the teacher with the most advanced group of students gave the second lowest set of TASP scores. These class-level results, both the teachers’ comments and the TASP scores, are interesting in that they support the notion of developmental and ability differences related to figurative competence. Moreover, they
demonstrate some level of inconsistency between what teachers may say in general and what they say about the progress of specific students, as well as inconsistency in some groups between teacher rating of student performance and students’ actual performance on a written test. This point about inconsistency will be discussed further in a later section.

In summary, although the statistical results of the study demonstrated limited significance relating ability and metaphor development as measured by the instruments used, student performance on the study tasks and teacher response to the intervention and to student results suggest that metaphor study is developmentally appropriate for verbally talented second graders according to definitions of developmental appropriateness (Bredekamp & Copple, 1997). Metaphor comprehension, interpretation, and production are not beyond the developmental abilities of these students, as demonstrated by the ability of some or many to respond adequately to the tasks, yet it is not too simple for the population, as demonstrated by the range of scores achieved and teachers’ confirmation of the challenge presented by the tasks. This generalization to verbally talented second graders is made with caution, because of the limited sample size and the small demonstration of statistical significance, but the results suggest at least that metaphor study has potential as a developmentally appropriate tool for learning in this population. Generalizations beyond the verbally talented among second graders would be inappropriate based upon these results, however, because the sample was entirely comprised of students from this group; the performance of students demonstrating a wider range of reading and verbal reasoning skills would need to be more carefully
explored before suggesting the efficacy of metaphor study for the larger population of second graders.

The Effects of Instruction

The discussion above centered on demonstration of the abilities comprising figurative competence as they were required by tasks within the instructional unit that served as the intervention. The results add to existing research around the developmental progression of abilities related to interpretation and production of metaphors by supporting predicted developmental patterns yet also demonstrating possibilities regarding advanced metaphor development in verbally advanced second graders. However, equally important in this study was the exploration of whether those abilities and the developmental progression they represent can be influenced by instruction around the structure and function of metaphor. This quasi-experimental aspect of the study reflected Vygotsky’s (1978, 1986) theories of the zone of proximal development and cognitive mediation, identifying instructional tools to serve as scaffolds for the interpretation of metaphor and exploring to what degree performance gains over time related to those tools could be observed.

In Levorato’s (1993) Model of Figurative Competence, the most advanced level of development is characterized by analysis and reflection upon figurative expressions, or metalinguistic competence, and incorporates the ability to process large amounts of language to analyze a figurative expression and the ability to use figurative language productively. These more advanced abilities were the target of the cognitive mediation tools in the study. The structure of metaphor was introduced and practiced through the tool of the Metaphor Chart, designed to aid students in approaching new figurative
expressions, and the Literature Web and subsequent discussions were intended to provide guidance for students in reflecting upon the purposes and effectiveness of metaphors they encountered. Moreover, activities in the instructional unit that used the Metaphor Chart as a stepping stone for supporting metaphor production were also intended to guide students toward the more advanced abilities suggested in the conceptual framework.

Previous research supported the notion that the use of these cognitive mediation tools would have a demonstrable effect on students' performance on the two instruments. One study of metaphor understanding in young gifted children demonstrated growth in comprehension related to an intervention utilizing analogical reasoning as a scaffold (Castillo, 1998), while other research has demonstrated growth gains in problem solving based on cognitive mediation with young gifted children in particular (Kanevsky, 1990). Moreover, research on curricular interventions grounded in the same framework as this study's intervention and utilizing the Literature Web has consistently demonstrated growth gains in literary analysis and interpretation attributable to the intervention (VanTassel-Baska, Johnson, Hughes, et al., 1996; VanTassel-Baska, Zuo, et al., in press).

However, none of the present study's results warranted definitive statements of similar conclusions. The analyses of pretest and posttest results within and between the treatment and comparison groups in this study were the primary sources for exploring whether the hypothesized instructional effects were achieved.

Results of the pretest-posttest analysis of the two major instruments of the study demonstrated statistical significance, but as described in the previous section, most of these results were in many respects inconclusive. Neither of the tests clearly demonstrated gains for the treatment group that could be attributed to the intervention,
and the small sample size of the comparison group in particular suggest a need for caution in interpreting statistical results that did appear. The MCT results were, indeed, statistically significant, with higher adjusted mean scores on the posttest for the treatment group than the comparison group, but the drop in scores from pretest to posttest within the treatment group is somewhat discouraging in terms of demonstrating growth gains. Indeed, as mentioned previously, the results raise questions about the ceiling of the instrument and the equivalence of the two forms as well as about the effects of the intervention.

The anomalous results for the treatment class receiving the most limited intervention were particularly interesting. It was encouraging to find that the drop in the treatment group's posttest scores was primarily the result of this one class; however, the drop within this class also far exceeded the small drop in scores experienced by the comparison group. It seems unlikely that an incomplete implementation of the intervention could have so strong a negative effect on students that it would outweigh even the experience of having no exposure to the intervention; variance resulting from the small sample size is a more likely explanation. However, the possibility does exist that the decrease in scores resulted from the influence of an ineffective teacher (Sanders, 2001); or conversely, there is the possibility that the stronger scores from other classes reflected more the effectiveness of their teachers than any effects of the intervention.

The class-specific analysis also revealed that the two treatment classes remaining in the analysis both demonstrated high mean scores on the pretest and little change between pretest and posttest. As with the overall analysis, the class-specific analysis also yielded a problem with ceiling effect on the instrument.
The results for the LAIT were more promising for the treatment group, with statistically significant pre-post growth gains, but the statistical results do not support attributing that growth to the intervention. None of the analyses comparing results for the two groups on the LAIT achieved statistical significance; therefore, despite gains that looked consistently larger for the treatment group at a descriptive level, the conclusion cannot be made that the growth was the result of participation in the intervention.

As with the MCT results, several possible explanations for the nonsignificant results of the LAIT analysis exist. Because more treatment group scores were available for the LAIT, the difference in sample size between the treatment and comparison groups was much larger in this analysis, and a few strong individual results in the comparison group had a particularly strong effect on mean scores. The activities of the intervention focused primarily on skills related to items 2 and 3 on the LAIT, but the rubric made only small gains on any one item possible; thus, it may be that the range of possible scores on the most relevant items was too limited to demonstrate sufficient change to support statistical significance. Although there was a span of several months between pretest and posttest, there is also the possibility of a test effect related to the items, the notion that the experience of taking the pretest may have influenced posttest results for both groups.

A further possible variable influencing the outcome of the analysis of the two instruments is the issue of fidelity of implementation. In keeping with standards for quasi-experimental design, efforts were made to ensure that the various classes and their experiences of the intervention were as similar as possible, despite the impossibility of random assignment. All of the treatment teachers experienced similar preparation for the study and were given the same guidelines for implementation, yet teacher self-reports and
observations by the researcher indicated some variability important to consider given the small teacher sample.

Possible results related to fidelity of implementation were most clear with regard to the teacher of Class 4, who implemented only a few lessons of the intervention. The large drop in MCT scores for this class from pretest to posttest and the confirmation that this teacher did not utilize the cognitive mediation tool of the Metaphor Chart are suggestive, especially given the strong connections between the items of the MCT and the structure of the Metaphor Chart. Although the score decrease is especially anomalous given that the comparison group’s scores remained nearly constant, Class 4’s results nevertheless offer some support for the inference that the results in treatment Classes 1 and 2 may be attributable to the intervention. Nevertheless, Class 4’s results on the LAIT did not show this same pattern of difference from the other classes; indeed, Class 4 not only showed gains but demonstrated higher gains than two of the other four classes, at a descriptive level. Interestingly, though, Class 4’s teacher was one of two who indicated use of the Literature Web throughout the year instead of only in conjunction with the intervention. As the skills required by the LAIT are intended to be supported by the Literature Web, this finding suggests the potential of an important relationship between a specific model for intervention and student gains.

On the other hand, the class taught by the teacher whose performance in an observed lesson raised concern did not demonstrate any clear patterns suggesting a relationship between fidelity of implementation and results. Class 3’s MCT results are difficult to judge, because the pretest was not administered until more than halfway through the intervention, but the posttest scores were not demonstrably lower than those

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
of other treatment classes. A potential for practice effect exists between the two administrations of the test in this group, but such an effect is impossible to judge from the data collected. Similarly, Class 3’s LAIT results fell in the middle of the group of treatment classes both in terms of posttest performance and gain scores. Thus, conclusions related to fidelity of implementation in this case are not so clear.

The student products collected during the course of the study were primarily collected to support investigation of the previous topic, related to the developmental appropriateness of the intervention for verbally talented second grade students. Because individual results were not systematically collected and analyzed across specific time frames of the intervention, conclusions cannot be drawn from the student products related to the effectiveness of the intervention. However, one suggestive point may be raised based on an analysis of two sets of student metaphors drawn from the same class at different points in the year, one set from early in the school year and the other from halfway through the intervention. The tree comparisons and half of the sun comparisons came from this class, and a review of specific statements and the categorization of these statements according to the metaphor characteristics studied demonstrated greater sophistication in the sun comparisons than the tree statements. Again, definitive attribution of these results to the intervention would be inappropriate based on the context of data collection and the possible influence of a maturity variable, but nevertheless the differences between the two sets of comparisons suggests that further exploration of the effects of the intervention on metaphor production is warranted.

In conclusion, the study results related to the effects of instruction on metaphor development in verbally talented second graders fail to demonstrate treatment effect. Yet
they also are to a large extent inconclusive and suggestive of the need for further exploration. Statistical significance favoring the treatment group on the MCT supports to some degree attribution of performance to the intervention, although caution is needed because of indications of instability of the instrument. Statistically significant pre-post growth gains on the LAIT for the treatment group suggest the possibility of some effective influence of the intervention, although the comparative results do not support that conclusion. In both cases, the small sample size of the comparison group may have influenced results and also prevents definitive generalizations or conclusions. Implementation differences among teachers in the treatment group also present possibilities for explaining the limited results but are themselves too limited to be conclusive.

Although all of the foregoing discussion has offered a variety of possible explanations for the inconclusive results of the study, another logical explanation is simply that the intervention itself did not have a measurable effect on student performance on tasks of figurative competence. Despite the suggestions regarding the influence of sample size, teacher effects, test effects, and other variables, this most fundamental of conclusions that the intervention was ineffective must also be considered. Moreover, because the procedures of the study involved real class time in the school setting, suggestions for further investigation related to the intervention must weigh the potential benefits of learning more about its possible effects with the potential danger of wasting instructional time with tasks that do not have an influence on student learning.
**Conclusion**

This study was intended to explore the notion of figurative competence as a developmental progression, with an eye to exploring developmental patterns related to metaphor comprehension and production and how those patterns emerge within the population of verbally talented second grade students. The study also investigated the effects of a specific instructional intervention on student performance of tasks assessing figurative competence.

The developmental patterns detected from the study findings with regard to student ability to comprehend, interpret, and produce metaphors were in line with predicted patterns and in some instances exceeded predicted patterns, which was an expected result based on the advanced verbal ability level of students in the sample. However, within the sample, only limited indications appeared of a predictive relationship between verbal ability as measured by a standardized test of verbal reasoning and figurative competence as measured by the study instruments. The possibility that these results were confounded by a ceiling on one or more of the measures or by instability of the study instruments exists and suggests the need for further investigation of the relationship between measures of verbal ability and figurative competence.

The findings related to the effects of instruction on figurative competence are also inconclusive. Some statistically significant treatment effects emerged, but primarily on a test whose overall results suggested a problem with ceiling effect on the instrument. Other growth gains also appeared but were not attributable to the intervention. Again, limitations created by the instruments and the sample size may have affected the results; the suggestions of possible effects, therefore, bear further investigation. Moreover, an
area of focus emerging from the progress of the study related to variability among teachers, in terms of their implementation, offered interesting possibilities for exploring relationships between teacher variables and student results. As the sample of teachers was extremely small and the relevant data not all systematically collected, no generalizable results emerged related to this study particularly; however, the tentative findings also suggest directions for further study. Nevertheless, the lack of treatment effect also suggests the possibility that the intervention itself will be ineffective even if the other limitations are overcome, and such a consideration must be addressed in terms of planning any future studies around it.

Based on the limitations created by sample size, instrumentation issues, and variable implementation, as well as the inconclusive findings related to each major area of focus, results from this study are not sufficiently strong to warrant definitive conclusions or generalizations beyond the sample. However, the qualitative and quantitative findings do offer a reasonable indication that metaphor study is developmentally appropriate for verbally talented second grade students in similar contexts to the students in the sample, and thus the potential for further investigation of instructional effects in the school context exists. Implications for such further investigations and tentative implications for practice are presented in the following section.

Implications for Research and Practice

Implications for Research

This study demonstrated few findings of statistical or educational significance regarding the effects of instruction on metaphor development. However, the study may be
viewed as an exploratory one, a preliminary investigation of the key questions, with several directions for additional research suggested by the findings.

First, the instruments utilized in the study for measurement of figurative competence require further analysis. The MCT, although designed according to recommendations regarding assessment of metaphor comprehension and piloted with a relevant sample, yielded troubling results suggesting a ceiling effect and weak equivalence between forms. Consequently, refinement of the instrument and subsequent piloting would be important steps to take before using it in further intervention research. Likewise, although the quantitative results of the LAIT were not so anomalous, the qualitative investigation of the item requiring metaphor interpretation also revealed a potential lack of equivalence between forms that would require attention if further investigation of that aspect were planned.

Although significant results related to treatment effects were minimal, the small sample size and the instrument issues addressed above suggest that the question of treatment effect has not been definitively answered, and observed teacher and student response to the intervention would suggest that it bears possibility for effect. Thus, additional research similar to this study is a worthwhile direction for further investigation. Moreover, additional research of this sort employing a larger number of classes would not only provide a larger student sample but would also allow for more investigation of the teacher variable than was possible in this study.

The teacher variable offers several interesting directions for further investigation. The use of the TASP by teachers as an effective instrument for assessing students’ figurative competence remains questionable, given the limited variability of scores
teachers assigned in this study and some inconsistency between commentary on student performance and rating of student performance. Refinement of the instrument, potentially including the incorporation of repeated use within one implementation of the intervention, might reveal additional results regarding student development as well as offering clearer indications of teacher understanding and use of the intervention and its underlying concepts.

Perhaps more promising as a direction for research is the issue of teacher attitude and its relationship to changes in student learning. Recent research efforts into teacher attitudes, teacher efficacy, and student achievement have demonstrated the important finding that changes in teacher attitudes and beliefs about student learning and the effectiveness of teaching practices occur not after professional development alone, but rather after the demonstration of changes in student learning following changes in classroom practice (Guskey, 2000). In this study, anecdotal evidence suggested some indication of a relationship between teacher attitudes in Classes 2 and 5 in particular, as demonstrated in informal conversation and commentary on the intervention, and student results. Future research efforts could explore teacher attitudes at various points within a similar implementation design to this one, collecting attitude data before and after the professional development workshop, just prior to implementation, during implementation, and at the conclusion; these data could then be examined for changes and for relationship to changes in student performance.

The qualitative portion of the study cast a wide net across many components of metaphor production and a component of comprehension and interpretation that was not included in the quantitative portion. The results of this portion of the study offered
interesting illustrations of predicted patterns of development in the students in the sample; each of the components investigated could in itself represent questions for further investigation. Moreover, as each of the patterns explored was drawn from existing quantitative research findings, each of them could also form the basis for further quantitative as well as qualitative investigation related to developmental patterns and intervention effects.

The variable of ability and its relationship to the development of figurative competence and to instructional effects could only be explored to a limited degree in this study because of the selection of the sample and the possible compression of the range of scores on the ability instrument used for comparison. However, the issue of accelerated developmental patterns in high ability students and the related possibilities of more powerful cognitive mediation effects (Kanevsky, 1990; Vygotsky, 1986) remains as an unanswered question with regard to metaphor development. Although the investigation of this issue in a school-based research setting as in this study is unlikely, given the demands for developing fluency and comprehension in primary reading classes, the issue could nevertheless form the basis of additional research in a more laboratory-based setting.

The issue of the setting for future research around the topics of the study is an important one to consider. As noted in the discussions of design and delimitations in Chapters I and III, this study specifically sought to investigate the effects of the intervention in existing classroom settings – the "real world" of schools – rather than in a lab school or enrichment program setting. The purpose of this design was to make findings of the study relevant to teachers in the most common settings for students of
strong verbal abilities in the early grades, which is in a regular language arts classroom. However, conducting research in the "real world" of schools runs the risk of many pitfalls, including teacher and student attrition as well as the interference of other demands on teachers' time and attention. The progress and results of this study were affected by many of these obstacles, including a low response of teacher volunteers despite a large number of teachers trained; a limited implementation by one treatment teacher because of timing and other demands on instructional time; and loss of pretest data in the treatment group because of misplaced papers. More significantly, although extensive written and oral communication regarding the expectations for participation was provided, a limited understanding of the research process and the controls it requires and perhaps an insufficient degree of follow-up communication resulted in less consistency over the study time frame than intended and the loss of an entire comparison class near the conclusion of data collection. These experiences are instructive for any researcher planning school-based studies, especially with regard to the need for clear and careful communication as well as strong control over the data collection process.

Implications for Practice

In terms of implications for practice, the student products and the results of the tests for both the treatment and comparison groups demonstrated that second grade students with advanced verbal skills are capable of understanding and producing metaphors, to varying degrees. The student products demonstrating complex, elaborated metaphors as well as those demonstrating simpler but nevertheless intentional metaphors support the notion that children at this age are capable of producing metaphors. Moreover, the analyses of metaphor demonstrated throughout the tests and the teacher
responses to the activities suggest that these students were certainly capable of understanding the structure of a metaphor, not just recognizing its format. Consequently, the study suggests that incorporation of the structure of figurative language into language arts study at the primary level can provide students with an opportunity to explore literature more deeply and to develop their expressive writing abilities.

Although the analyses relating student ability to performance on the study instruments yielded only weak results, the finding that some of these results were statistically significant and the commentary by teachers regarding different students' response to the intervention bear implications for classroom practice. According to the teacher commentary, different students responded to the activities to different degrees, with most students feeling challenged and some able to accomplish more than others, though in general students were engaged with the material. This reflects the understanding that within the population of gifted students, one finds considerable variability in terms of ability in general and in specific areas, and that the degree of challenge beyond what the "regular" classroom offers needs itself to be differentiated within the gifted population. Moreover, this differentiation needs to take into account the different areas within the language arts at which students may be particularly talented or in need of greater challenge, with appropriate response in terms of the different types of tasks with which students are presented across the verbal domain. Responses from teachers regarding the challenge the intervention presented even to their more advanced students and the teachers' own surprise at some of the results support recommendations for providing challenging classroom experiences appropriate for young students with advanced talents in the verbal areas.
The study findings also suggest an implication related to fidelity of implementation of curriculum. In the study, the teacher who implemented the fewest lessons of the unit had the class with the lowest posttest results on one instrument and nearly the lowest on the other. Although the instrumentation issues and sample size limit the conclusions to be drawn from the class-level analyses, there are nevertheless strong indications to suggest that an intervention implemented as intended has greater potential to enhance learning than an intervention implemented to a more limited degree.

Education is plagued with the problem of halfhearted attempts to implement innovations, with many projects abandoned before they have been sufficiently tried; the issue of fidelity of implementation related to any innovation or specific intervention is an important one to consider both in research and in classroom practice.

The study also carries implications related to practices in professional development. Anecdotal results suggested a clear connection between teacher attitudes and student results, which is supported by the professional development literature (Guskey, 2000). Moreover, the teacher who demonstrated the most positive change in attitude not only saw results in her students related to her own teaching, but also sought and received the chance to observe her students' response to another teacher's demonstration of the intervention early in the study, while she herself was still tentative about its potential and appropriateness for her group. This finding, although not generalizable because it comes from the experience of only one teacher, nevertheless supports the notion that professional development should include not only workshops and monitoring, but also opportunities for teachers to observe interventions in action with their own students to support attitude change.
Summary

Metaphor is an important component of speech and writing, a support for new understandings and for new explorations of familiar and unfamiliar words and their referents. The development of figurative competence is an ongoing process in children as they learn and explore words, finding literal and figurative meanings and experimenting with the use of those meanings in creating imagery. This study was intended to explore that developmental process, most particularly in children whose developmental patterns in other areas tend to be accelerated, to discover if figurative competence bears a significant relationship with other verbal ability and whether the developmental process around the understanding and production of metaphor could be advanced through specific albeit short-term instructional intervention.

The findings did demonstrate patterns of performance in students reflecting predictions for their age and beyond, all suggesting that second graders of advanced verbal abilities have some capacity to understand, interpret, and produce metaphorical comparisons. The intervention findings of the study, however, provided only tentative suggestions of significant relationships between the key variables, uncovering issues related to instrumentation and implementation that may or may not have affected the results and raising questions for further investigation. The findings also hinted at important connections between the teacher as variable and student results, again providing directions for future research. Furthermore, the findings could also be interpreted to support the conclusion that the intervention was ineffective and should not be investigated further, although the many limitations and questions would also make this conclusion somewhat precipitate.
In truth, the study raised many more questions than it answered, and should be viewed as an exploratory investigation of the combination of concepts included. The results provide a foundation for further investigation, including the early stages of development of an instrument to measure young children's comprehension of metaphors in context. The intervention itself represents a potentially promising curriculum for supporting metaphor development or at least for providing verbally talented second graders with the opportunity to encounter and explore the structure of metaphor and its place in the language they speak, hear, read, and write. These contributions are important in terms of the directions they provide for further research and practice in the areas of early childhood gifted education, primary language arts education, and the study of figurative competence as an aspect of linguistic learning and development in children.
REFERENCES


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Sanders, W. L. (2001, June). The impact of teaching on student achievement. Presentation at the William and Mary School Leadership Institute, Williamsburg, VA.


APPENDIX A

Demographic Data on Schools

<table>
<thead>
<tr>
<th>School</th>
<th>Total Enrolled</th>
<th>Average Class Size</th>
<th>% F/R Meals</th>
<th>% Racial/Ethnic Composition</th>
<th>Meeting Local CRT Standard in Reading/Math at gr 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Afr.</td>
<td>Amer.</td>
</tr>
<tr>
<td>School 1*</td>
<td>460</td>
<td>23.3</td>
<td>36.1</td>
<td>32.8</td>
<td>.4</td>
</tr>
<tr>
<td>School 2</td>
<td>500</td>
<td>24.1</td>
<td>2.6</td>
<td>3.8</td>
<td>.2</td>
</tr>
<tr>
<td>School 3</td>
<td>895</td>
<td>27.3</td>
<td>19.1</td>
<td>21.3</td>
<td>.4</td>
</tr>
<tr>
<td>School 4</td>
<td>634</td>
<td>25.7</td>
<td>3.5</td>
<td>3.8</td>
<td>.3</td>
</tr>
<tr>
<td>School 5</td>
<td>673</td>
<td>24.6</td>
<td>38.8</td>
<td>30.5</td>
<td>.1</td>
</tr>
<tr>
<td>School 6</td>
<td>594</td>
<td>25.6</td>
<td>43.1</td>
<td>36.7</td>
<td>.2</td>
</tr>
</tbody>
</table>

* School 1 participants in study drawn from gifted magnet population within the school; separate magnet program demographic data not provided.

Demographic data drawn from 1999-2000 school year reports.
APPENDIX B

Metaphor Comprehension Test

FORM A

Directions:
Read the poems and circle the letter for the best answer to each question.
Note: All poems may be found in the following texts, as indicated:
    Boston: Little, Brown.

“Rags” by Judith Thurman (*A Year Full of Poems*, p. 106)

1. What is the poem mostly about?
   a. a doorknob
   b. the night wind
   c. rags
   d. October

2. Which of these sentences about a doorknob is important for understanding this poem?
   a. You can open a door by turning a doorknob.
   b. Sometimes doorknobs have keyholes in them.
   c. Doorknobs are round and sometimes shiny.
   d. If a door is locked, the doorknob will not turn.

3. What does the poem tell us about clouds?
   a. Some clouds are white.
   b. Some clouds look like thin strips of cloth.
   c. Some clouds are fluffy like cotton balls.
   d. Clouds are a sign that it will rain soon.

4. What does the poem tell us the moon is like?
   a. the sun
   b. rags
   c. October
   d. a doorknob

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
“Snowy Benches” by Aileen Fisher (Knock at a Star, p. 66)

5. What thing in the poem is like a person?
   a. snow
   b. a bench
   c. winter
   d. loneliness

6. Which of these sentences has the same main idea as the poem?
   a. Lonely people like to go to the park.
   b. In winter benches can get covered with snow.
   c. In winter there are not many people visiting the park.
   d. It snows in the wintertime.

“If You Catch a Firefly” by Lilian Moore (A Jar of Tiny Stars, p. 48)

7. What does the poet say a firefly is like?
   a. a bee
   b. a jar
   c. a light bulb
   d. a star

8. What is the most important thing the poet is saying about fireflies?
   a. They fly.
   b. You can catch them at night.
   c. They are bright.
   d. You don’t see them much in the daytime.

9. Which of these sentences about stars is important for understanding the poem?
   a. The sun is a star.
   b. Stars are twinkly and bright.
   c. You can see many stars at night.
   d. Stars are made of hot gases.

“Rain Poem” by Elizabeth Coatsworth (Sing a Song of Popcorn, p. 29)

10. What came in the window in this poem?
    a. a mouse
    b. the rain
c. tracks

d. the wind

11. What description of the rain is important in this poem?
a. Sometimes thunder and lightning come with rain.
b. Rain helps plants to grow.
c. Rain can make soft noises on the ground and roof when it falls.
d. Umbrellas and raincoats are good to have on rainy days.

12. What are the "tracks across the sill"?
a. raindrops
b. mouse footprints
c. train tracks
d. dirt

13. Which of these sentences about mice is important for understanding the poem?
a. Mice eat cheese.
b. Mice have tails.
c. Cats chase mice.
d. Mice are not loud.

Answer Key (Form A):
1. b
2. c
3. b
4. d
5. b
6. c
7. d
8. c
9. b
10. b
11. c
12. a
13. d
FORM B

Directions:
Read each poem and circle the letter for the best answer to each question.

from "The Wind" by James Stephens (Knock at a Star, p. 100)
1. What does the poem show the wind is like?
   a. a fan
   b. a person
   c. a tornado
   d. a tree
2. What is the most important thing the poet is trying to show about the wind?
   a. The wind is noisy.
   b. The wind is cold.
   c. The wind is angry.
   d. The wind is warm.
3. Which sentence has the same main idea as the poem?
   a. The wind is a person who can shout and whistle on his fingers and kick.
   b. The leaves were withered and the branches were bare so it must be autumn.
   c. The wind started blowing with lots of noise and carrying the leaves through the air.
   d. Shouting, whistling, kicking, and thumping can make a lot of noise.

"safety pin" by Valerie Worth (A Jar of Tiny Stars, p. 63)
4. What is the poem mostly about?
   a. a shrimp
   b. a fish
   c. an eye
   d. a safety pin
5. When does a safety pin seem to be sleeping?
   a. when it is closed
   b. when it is in a box
   c. when it is open
   d. when you prick your finger

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
6. Which of these sentences about a small fish is important for understanding the poem?
   a. It lives in the water.
   b. It has scales.
   c. It uses fins to swim.
   d. It has a long oval shape.

7. What is the tail in the poem?
   a. shrimp's tail
   b. sharp part of a pin
   c. fish's tail
   d. a silver image

   from “Firefly” by Li Po (Sing a Song of Popcorn, p. 129)

8. Who is the speaker in the poem talking to?
   a. another person
   b. the moon
   c. himself
   d. a firefly

9. What is something about stars that is important in this poem?
   a. You can see many stars at night.
   b. Stars are made of hot gases.
   c. Stars are bright and twinkly.
   d. The sun is a star.

   “What is The Sun?” by Wes Magee (A Year Full of Poems, p. 64)

10. Which thing does the poet NOT say is like the sun?
    a. paper
    b. coin
    c. ball
    d. bottle top

11. What things about the sun is the poet talking most about?
    a. the sun's shape and heat
    b. the sun's shape and color
    c. the sun's heat and color
    d. the sun's heat and movement
12. What does the poet think is like a calm sea?
   a. a quiet river
   b. a puddle
   c. the sky
   d. the sun

13. Which sentence about the sea is important for understanding the poem?
   a. Many animals live in it.
   b. It is large and blue.
   c. You can swim in the sea.
   d. Sea water is salty.

Answer Key (Form B):
1. b
2. a
3. c
4. d
5. a
6. d
7. b
8. d
9. c
10. a
11. b
12. c
13. b
APPENDIX C

Literary Analysis and Interpretation Test

FORM A

Fog

The fog comes in
on little cat feet.

It sits looking
over harbor and city
on silent haunches
and then moves on.

— Carl Sandburg

Read "Fog" and answer the questions.

1. What do you think is the main idea of this poem? Use a sentence or two to tell what you think.

2. What do you think the poet meant by the words, "The fog comes in on little cat feet"?

3. a. What is the main thing the poem is about?

   b. Look at your answer to 3a. When you read the poem, did it make you think about that thing in a new way? Explain how.

4. Make up a new title for the poem. Tell why you think your new title would be a good one.
FORM B

April Rain Song (from Sing a Song of Popcorn, p. 20)

Read “April Rain Song” and answer the questions.

1. What do you think is the main idea of this poem? Use a sentence or two to tell what you think.

2. What do you think the poet meant by the words, “Let the rain sing you a lullaby”?

3. a. What is the main thing the poem is about?

   b. Look at your answer to 3a. When you read the poem, did it make you think about that thing in a new way? Explain how.

4. Make up a new title for this poem. Tell why you think your new title would be a good one.
APPENDIX D

LAIT Scoring Rubric

Form A scoring and sample responses – “Fog”

1. What do you think is the main idea of this poem? Use a sentence or two to tell what you think.

   0 Provides no response or a response inappropriate to the task demand. (e.g., I don’t know)

   2 limited, vague, inaccurate, confusing, only quotes from reading
   Samples:
   One important idea is it comes in on little cat feet.
   It’s about fog.

   4 accurate but literal, limited; limited demonstration of metaphor comprehension
   Samples:
   How fog comes and goes.
   It’s trying to tell you what fog sees.
   That the fog is silent.
   That fog is like cats.

   6 meets expectations: demonstrates understanding of the central metaphor; must include reference to topic, vehicle, and ground
   Samples:
   The important idea is how fog comes in quietly like a cat.
   That fog is gray and sneaky like a cat.

   8 exceeds expectations: insightful response that offers fluent, substantial support
   (everything from 6 score plus extra elaboration)

2. Use your own words to explain what you think the author means by the words “The fog comes in on little cat feet.”

   0 Provides no response or a response inappropriate to the task demand.

   2 limited, vague, inaccurate; rewording or restating only
   Samples:
   That it comes in on cats feet.
   Because it has cat feet.
   Because cats can see through fog.

   4 accurate but limited response; recognizes ground of comparison but does not explicitly explain
   Samples:
   that he was trying to say fog is low
   it means how the fog moves
   it means it comes quietly

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
6 *meets expectations*: demonstrates understanding of the metaphor and includes topic, vehicle, and ground in response

Samples:

*That fog comes in slowly like a cat's feet.*

*Fog comes in quietly and when a cat walks you don't hear a sound*

*How quietly it comes and how gently because when cats walk it is usually gentle and quiet.*

8 *exceeds expectations*: insightful response with substantial support (everything from 6 score plus extra elaboration)

3. a. What is the main thing the poem is about? b. When you read the poem, did it make you think about that thing in a new way? Explain how.

0 Provides no response or a response inappropriate to the task demand, or if wrong response given to 3a with no explanation in 3b (e.g., “cat” with no additional explanation)

2 limited, vague; or if wrong response to 3a given but connection to the topic in 3b (reversed metaphor).

Samples:

* a. fog b. no
  a. fog b. because it made me think about fog
  a. cat b. it made me think how cats are like fog

4 correct response to 3a with literal, limited explanation in 3b

Samples:

* a. fog b. because it said fog is like a cat
  a. fog b. fog is quiet

6 *meets expectations*: correct response to 3a; explanation in 3b must demonstrate recognition of the poem’s metaphor (including reference to vehicle and ground)

Samples:

* a. fog b. it showed how fog moves quietly like a cat
  a. fog b. no because I already thought fog was quiet and gray like a cat

8 *exceeds expectations*: (everything from 6 score plus extra elaboration)

4. Make up a new title for the poem. Tell why you think your new title would be a good one.

0 Provides no response or a response inappropriate to the task demand.

2 limited, vague, inaccurate, or title given without explanation

Samples:

*Fog and the Cat
I Like Fog
The Cat because it's about a cat out in the fog*
4 literal response with an attempt to support
Samples:
The Fog Moves In and Out because it mostly talks about the way the fog moves
Silent Fog because fog is quiet
Low Cloud because it would be another name for fog

6 meets expectations: meaningful title given with appropriate support; addresses topic, vehicle, and ground
Samples
Silent Fog because it says the fog moves like a cat which is silent
The Cat Like Fog because it shows fog is quiet like a cat
The Gray Cat Fog because the poem is about how the fog is like a gray cat

8 exceeds expectations: insightful title given with multiple applications from text (everything from 6 score plus extra elaboration)

Form B scoring and sample responses – “April Rain Song”
1. What do you think is the main idea of this poem? Use a sentence or two to tell what you think.

0 Provides no response or a response inappropriate to the task demand. (e.g., I don’t know)

2 limited, vague, inaccurate, confusing, only quotes from reading
Samples:
It's about rain.
Let the rain kiss you.
It's about rain in April.
It's a rain song.

4 accurate but literal, limited; limited demonstration of metaphor comprehension
Samples: 
It's telling how the rain can sound.
What the rain does.
It's the sounds of the rain..

6 meets expectations: demonstrates understanding of the central metaphor; must include reference to topic, vehicle, and ground
Samples: 
It's about how rain sounds like a song.
That rain can be like a lullaby in how it sounds..

8 exceeds expectations: insightful response that offers fluent, substantial support (everything from 6 score plus extra elaboration)
2. Use your own words to explain what you think the author means by the words “Let the rain sing you a lullaby.”

0 Provides no response or a response inappropriate to the task demand.

2 limited, vague, inaccurate; rewording or restating only
Samples:
Because rain sings.
It’s a lullaby.
It’s a song.

4 accurate but limited response; recognizes ground of comparison but does not explicitly explain
Samples:
It means the sounds of the rain.
How the rain sounds.
The rain sounds good.

6 meets expectations: demonstrates understanding of the metaphor and includes topic, vehicle, and ground in response
Samples:
The rain makes quiet sounds like a lullaby.
It’s like the rain is singing to you when it’s raining.

8 exceeds expectations: insightful response with substantial support (everything from 6 score plus extra elaboration)

3. a. What is the main thing the poem is about? b. When you read the poem, did it make you think about that thing in a new way? Explain how.

0 Provides no response or a response inappropriate to the task demand, or if wrong response given to 3a with no explanation in 3b (e.g., “song” with no additional explanation)

2 limited, vague; or if wrong response to 3a given but connection to the topic in 3b (reversed metaphor).
Samples:
b. rain  b. no
b. rain  b. because it made me think about rain
a. song  b. it made me think how a song is like rain

4 correct response to 3a with literal, limited explanation in 3b
Samples:
b. rain  b. because it said the rain is like a song
a. rain sounds  b. rain is singing
6  meets expectations: correct response to 3a; explanation in 3b must demonstrate recognition of the poem's metaphor (including reference to vehicle and ground)
   Samples:
   b. a rain song b. it showed how rain makes pretty sounds like a song
   a. rain  b. because when it rains it sounds like someone is singing to you

8  exceeds expectations: (everything from 6 score plus extra elaboration)

4. Make up a new title for the poem. Tell why you think your new title would be a good one.

0  Provides no response or a response inappropriate to the task demand.

2  limited, vague, inaccurate, or title given without explanation
   Samples:
   Rain Song
   Rain and April
   Singing in the Rain
   Let the Rain Kiss You

4  literal response with an attempt to support
   Samples:
   Rain in April because the poem talks about what the rain sounds like in April
   Singing Rain because the rain sings a lullaby

6  meets expectations: meaningful title given with appropriate support; addresses topic, vehicle, and ground
   Samples
   Raining Lullaby because it tells how rain can sounds quiet like a lullaby
   April Singing because it rains in April and it sounds like singing when it rains

8  exceeds expectations: insightful title given with multiple applications from text
   (everything from 6 score plus extra elaboration)
APPENDIX E

Teacher Assessment of Student Performance

Please complete the form for each student in the unit based on your assessment of their participation in discussions and written activities. An explanation and possible sample activity are given for each item as explanation.

Student ID:

Scale:
1 = did not demonstrate or was unable to demonstrate; developmentally unready
2 = made clear attempts to demonstrate, but generally with weak responses or responses that he/she could not explain further
3 = consistently demonstrated at a satisfactory level through at least the second half of the unit (meets expectation of item)
4 = consistently demonstrated advanced figurative competence (exceeds expectation of item)
UA = unable to assess

The student...
1. showed the ability to identify meanings of words in figurative expressions based on context.
   e.g., explained how other words, phrases, or pictures helped to show the correct meaning to be used to understand a given word
   (example activity: in discussion of poems or stories)  
   1  2  3  4  UA

2. showed the ability to explain multiple meanings of words and why a given meaning was appropriate for understanding a figurative expression.
   e.g., identified multiple meanings of a given word, explained why one meaning and not a different meaning was the appropriate one to use in interpreting a given expression
   (example activity: in discussion of poems or stories, especially "Foolish Questions")  
   1  2  3  4  UA

3. showed the ability to recognize that given figurative expressions were not supposed to be interpreted literally.
   e.g., explained that a metaphorical statement could not really be "true" and why; or recognized and explained the humor or imagery in a figurative expression
   (example activity: in discussion of poems or stories, use of literature web esp. key words and images)  
   1  2  3  4  UA
4. showed the ability to identify the topic of a figurative expression
e.g., was able to identify which object in a metaphorical statement or longer piece was what the figurative expression was actually about
(example activity: responses to first column of comparison chart)

5. showed the ability to identify the vehicle of a figurative expression
e.g., was able to identify which object in a metaphorical statement or longer piece was what the topic was being compared to
(example activity: responses to second column of comparison chart)

6. showed the ability to identify the relevant grounds of comparison in a figurative expression
e.g., listed the various characteristics of a topic and vehicle and identified the characteristics held in common by both that were being emphasized in a given metaphorical expression
(example activity: responses to third column of comparison chart)

7. showed the ability to infer author purpose regarding the topic of a figurative expression
e.g., was able to explain what characteristics of a given metaphor topic a poet especially wanted to emphasize by using a specific figurative comparison
(example activity: in discussion of poems or stories, especially after comparison chart; also ideas bubble of literature web)

8. showed the ability to process large amounts of text at once to analyze meaning of figurative expressions
e.g., gave an appropriate interpretation of a metaphorical sentence quoted from a poem or stated the main idea of a longer metaphorical piece
(example activity: in discussion of poems or stories; key words or ideas bubbles of literature web)

9. showed the ability to produce figurative comparisons based on specific characteristics of the topic of the comparison and what it was compared to.
e.g., wrote a poem or metaphorical statement highlighting specific characteristics of a topic through comparison to an appropriate vehicle
(example activity: sun poems)

10. showed the ability to use figurative language spontaneously to explain or emphasize something about a given topic
e.g., used a metaphor or analogy to demonstrate understanding of a new word or expression; used an unusual metaphorical comparison to emphasize particular features of a topic
(spontaneous student behavior during unit instruction or otherwise in school day)

Please add any comments on back to elaborate.
## APPENDIX F
### Classroom Observation Form

<table>
<thead>
<tr>
<th>Name of Observer ___________________________</th>
<th>Minutes of Observation ____________</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time __________ Date __________ Lesson Observed ____________</td>
<td></td>
</tr>
<tr>
<td>Name of School ___________________________</td>
<td>Name of Teacher ____________________</td>
</tr>
<tr>
<td>Number of Students __________ Grouping Model ____________________</td>
<td></td>
</tr>
</tbody>
</table>

### Area

#### The Teacher ...

<table>
<thead>
<tr>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. instructed/practiced literary analysis &amp; interpretation (literature web)</td>
</tr>
<tr>
<td>2. instructed/practiced word analysis (vocabulary web)</td>
</tr>
<tr>
<td>3. instructed/practiced persuasive writing (hamburger model)</td>
</tr>
<tr>
<td>4. instructed/practiced metaphor analysis (comparison chart)</td>
</tr>
<tr>
<td>5. engaged students in oral discussion of literature using high-level questions</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. instructed/practiced analogical reasoning</td>
</tr>
<tr>
<td>7. instructed/practiced student research</td>
</tr>
<tr>
<td>8. modeled/practiced higher level thinking&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>9. debated points of view</td>
</tr>
<tr>
<td>10. modeled/practiced metacognition&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. instructed/practiced concept mapping</td>
</tr>
<tr>
<td>12. emphasized “change” in instruction and assignments</td>
</tr>
<tr>
<td>13. instructed/applied the unit generalizations about change</td>
</tr>
<tr>
<td>14. emphasized other relevant key concepts, themes or ideas in instruction and assignments</td>
</tr>
<tr>
<td>15. encouraged/indicated interdisciplinary and/or real world connections</td>
</tr>
</tbody>
</table>

---

1. Higher-level thinking refers to strategies and skills such as making comparison and contrast; judging and evaluating situations, problems, or issues; generalizing from specific data to the abstract; synthesizing or summarizing information within or across the disciplines.

2. Metacognition refers to strategies and skills such as planning, monitoring, self-reflection or self-appraisal, and thinking about their own thinking.

Observer’s Signature _____________________________

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
APPENDIX G

Curriculum Framework and Lesson Outline

CONTENT GOALS

Goal 1: To develop analytical and interpretive skills in literature.

Students will be able to...

1. Describe what a selected literary passage means.
2. State an important idea of a reading.
3. Analyze similarities and differences in meaning among selected works of literature.
4. Create a title for a reading selection and provide a rationale to justify it.

Goal 2: To develop skills in identifying, analyzing, and using figurative language.

Students will be able to...

1. Recognize figurative expressions in text, including simile, metaphor, and personification.
2. Analyze a metaphorical expression for topic, object of comparison, and shared characteristics.
3. Use the forms of simile, metaphor, and personification to create figurative comparisons.

Goal 3: To develop persuasive writing skills.

Students will be able to...

1. Write a persuasive paragraph that includes a claim, reasons, and conclusion.
2. Revise and edit a piece of writing.

Goal 4: To develop linguistic competency.

Students will be able to...

1. Use context clues and analogies to discover word meanings.
2. Develop vocabulary skill commensurate with reading.
PROCESS GOAL

Goal 5: To develop analogical reasoning skills.

Students will be able to...

1. Solve verbal analogy problems.
2. Use analogies as support for understanding new words.
3. Recognize and analyze figurative language that appears in analogy form.

CONCEPT GOAL

Goal 6: To develop an understanding of the concept of change, especially changes related to language.

Students will be able to...

1. Understand that change is linked to time.
2. Analyze changes to determine whether they are positive or negative, natural or human in cause, and orderly or random.
3. Recognize the change process at work in a selection of literature.
4. Demonstrate changes in language over time.
5. Describe changes language can cause in human behavior and emotions.

Lesson Outline

1. Introduction and Pre-Assessment
2. The Concept of Change
3. Language and Change
4. Adjectives
5. The Unicorn and the Moon
6. Creating Imagery
7. Similes and Metaphor
8. Owl Moon
9. Persuasive Writing
10. Analogies
11. Understanding Words in Context
12. Haiku and the Seasons
13. Personification
14. Symbols
15. Wordless Books
16. *The Mysteries of Harris Burdick*
17. Concrete Poems
18. Project Presentations
19. Unit Wrap-Up and Post-Assessment
## APPENDIX H

Qualitative Analysis of Metaphor Production

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>my cat is as soft as a pillow</td>
<td>✓</td>
<td>A</td>
<td>A</td>
<td>touch - texture</td>
<td>even</td>
<td></td>
<td>v. high salience topic</td>
</tr>
<tr>
<td>my blood is as red as a red crayon</td>
<td>✓</td>
<td>F</td>
<td>A</td>
<td>visual - color</td>
<td>H-L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>water can be as cold as ice</td>
<td>✓</td>
<td>F</td>
<td>A</td>
<td>visual - color</td>
<td>L-H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the house was as big as an elephant</td>
<td>✓</td>
<td>F</td>
<td>A</td>
<td>visual - size</td>
<td>even</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the children were as quiet as a mouse</td>
<td>✓</td>
<td>A</td>
<td>A</td>
<td>sound - volume</td>
<td>L-H</td>
<td></td>
<td>common</td>
</tr>
<tr>
<td>the classroom was as colorful as a rainbow</td>
<td>✓</td>
<td>A</td>
<td>A</td>
<td>visual - color</td>
<td>L-H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a ball is as round as the moon</td>
<td>✓</td>
<td>F</td>
<td>A</td>
<td>visual - shape</td>
<td>even</td>
<td></td>
<td>v. high salience topic</td>
</tr>
<tr>
<td>the lion is as little as a baby</td>
<td>✓</td>
<td>F</td>
<td>A</td>
<td>visual - size</td>
<td>L-H</td>
<td></td>
<td>topic requires clarif.</td>
</tr>
<tr>
<td>the football player was as rough as a rock</td>
<td>✓</td>
<td>F</td>
<td>A</td>
<td>touch - texture</td>
<td>even</td>
<td></td>
<td>weak T-V pairing</td>
</tr>
<tr>
<td>the teacher roars as much as a lion</td>
<td>✓</td>
<td>A</td>
<td>R</td>
<td>behav.</td>
<td>L-H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mr Z’s shirt is as green as Cody’s shirt</td>
<td>✓</td>
<td>A</td>
<td>A</td>
<td>visual - color</td>
<td>even</td>
<td></td>
<td>high salience topic</td>
</tr>
<tr>
<td>the baby was as small as a teddy bear</td>
<td>✓</td>
<td>F</td>
<td>A</td>
<td>visual - size</td>
<td>even</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a soft mountain so tall reaching to grab the sugar</td>
<td>✓</td>
<td>A</td>
<td>R</td>
<td>behav.</td>
<td>L-H</td>
<td>maint.</td>
<td>embeds a second metaphor</td>
</tr>
<tr>
<td>snow to be so white</td>
<td>✓</td>
<td>A</td>
<td>A</td>
<td>visual - color</td>
<td>even</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rocks can be cool things they’re all different sizes just like lumpy balls</td>
<td>✓</td>
<td>A</td>
<td>A</td>
<td>visual - shape</td>
<td>even</td>
<td></td>
<td>elab. clarity</td>
</tr>
<tr>
<td>grass covers the mount fire blazing on the ground when the sun beams reach...</td>
<td>✓</td>
<td>A</td>
<td>A</td>
<td>visual - color</td>
<td>L-H</td>
<td>maint.</td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Trees are like a pillow connected to a sick. Trees give us shade and trees give us some shine. Trees is a playground and a home at the same time. Trees gives us food.</td>
<td>✓</td>
<td>F</td>
<td>A</td>
<td>visual-shape</td>
<td>H-L</td>
<td>lose</td>
<td></td>
</tr>
<tr>
<td>A tree is like a tall plant because a tree starts out like a plant. A tree can be small and a plant can be small.</td>
<td>✓</td>
<td>A</td>
<td>R</td>
<td>func.</td>
<td>L-H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The trees have leaves and telephone poles don't, but the telephone pole with all its clipings and nippings, looks a little modern, and the trees with all its bows blossoms every summer.</td>
<td>✓</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>weak T-V pairing</td>
</tr>
<tr>
<td>Trees can compare to anything. Like a king. Maybe a bell that doesn't ring.</td>
<td>✓</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>weak T-V pairing</td>
</tr>
<tr>
<td>Trees are different from brockly. Trees are big and hard. Brockly is small and soft.</td>
<td>✓</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tree are like a cloud. Because tree move when the wind is coming like a cloud. Tree and cloud can look alike and different.</td>
<td>✓</td>
<td>F</td>
<td>A</td>
<td>visual-movement</td>
<td>L-H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some trees look like monsters. I do know why. I used they looked so scary as my car went by.</td>
<td>✓</td>
<td>V</td>
<td>A</td>
<td>visual-shape</td>
<td></td>
<td></td>
<td>weak T-V pairing</td>
</tr>
<tr>
<td>Trees are very big. They are big like a pig. They are fat and wide like a pig.</td>
<td>✓</td>
<td>F</td>
<td>A</td>
<td>visual-size</td>
<td>even</td>
<td></td>
<td>weak T-V pairing</td>
</tr>
<tr>
<td>A tree is shaped like a streetlight, because of its tall height. A tree is like a streetlight, because it doesn't fit. A tree is like a streetlight.</td>
<td>✓</td>
<td>F</td>
<td>A</td>
<td>visual-shape, size</td>
<td>H-H</td>
<td></td>
<td>T-V too close on ground</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>--------</td>
<td>-------------------</td>
<td>--------------</td>
<td>-----------------------------</td>
<td>--------</td>
<td>--------------</td>
<td>------------------</td>
</tr>
<tr>
<td>the sun is like a lamp, separated from earth by an invisible ramp. It is like a ball of light hanging in the air. It is like a sphere. But I wouldn’t touch I wouldn’t dare!</td>
<td>✓</td>
<td>A</td>
<td>A</td>
<td>visual-light/func.</td>
<td>H-H</td>
<td>lose</td>
<td></td>
</tr>
<tr>
<td>the sun is like a peice gold but you never touch or hold. the sun is first it stays until the moon takes birth</td>
<td>✓</td>
<td>A</td>
<td>A</td>
<td>visual-color, shape/cycle</td>
<td>even</td>
<td>lose</td>
<td>two metaphors</td>
</tr>
<tr>
<td>the sun is like a giant light bulb. It’s so burning hot. It pours over the earth like lava. Like so much burning fire.</td>
<td>✓</td>
<td>A</td>
<td>A/R</td>
<td>visual-light/move.</td>
<td>H-H</td>
<td>lose</td>
<td>two metaphors</td>
</tr>
<tr>
<td>The sun’s bright. The sun’s a globe of light. In the night you see her light in a different place.</td>
<td>✓</td>
<td>A</td>
<td>A</td>
<td>visual-shape</td>
<td>H-H</td>
<td>lose</td>
<td>T-V close on ground</td>
</tr>
<tr>
<td>The sun is like a burning stove.... It’s hot it’s fiery...</td>
<td>✓</td>
<td>A</td>
<td>A</td>
<td>touch-temp.</td>
<td>H-H</td>
<td>lose</td>
<td></td>
</tr>
<tr>
<td>The sun is like a ball of sring, weving it’s light through space like a never-ending pecie of thred. Nether old, nether yong, it’s light is an everlasting memery.</td>
<td>✓</td>
<td>A</td>
<td>R</td>
<td>cycle</td>
<td>L-H</td>
<td>maint</td>
<td>second metaphor elab. on first</td>
</tr>
<tr>
<td>The sun is like water, somtimes nice and bright. But somtimes hot and mean and draws famine in.</td>
<td>✓</td>
<td>A</td>
<td>A/R</td>
<td>visual-touch/phen</td>
<td>H-L</td>
<td>maint</td>
<td>unusual elab</td>
</tr>
<tr>
<td>The sun is like a globe of light. In some places she shines in the night.</td>
<td>✓</td>
<td>A</td>
<td>A</td>
<td>visual-shape</td>
<td>H-H</td>
<td>lose</td>
<td>T-V close on ground</td>
</tr>
<tr>
<td>The sun is like a ball of light. The sun is very bright. But if you catch it. It will fly away at night. Even though you know I’m wrong the sun is quite a fright.</td>
<td>✓</td>
<td>A</td>
<td>A</td>
<td>visual-shape</td>
<td>H-H</td>
<td>lose</td>
<td>attempt at addl elab but weak</td>
</tr>
<tr>
<td>The sun is like a star. Big bright and butiful waiting all day for moon to rise so he can go to sleep.</td>
<td>✓</td>
<td>A</td>
<td>R</td>
<td>behav</td>
<td>L-H</td>
<td>lose</td>
<td>first part literal/sec. met.</td>
</tr>
<tr>
<td>The sun is a golden ball shimmering in the sky, who would dare to look at well I know well it would not be I. The stars play with throwing it up really high. Then it goes to the other side and the moon gets thrown over again.</td>
<td>✓</td>
<td>A</td>
<td>A/R</td>
<td>visual-shape, color</td>
<td>L-H</td>
<td>maint</td>
<td>elab. adds strength</td>
</tr>
<tr>
<td>The sun is like a lemon drop. That’s very very big. It’s such a big ol’ ball of light, it acts like it’s shiny. It just</td>
<td>✓</td>
<td>A</td>
<td>A</td>
<td>visual-color, shape</td>
<td>even</td>
<td>lose</td>
<td></td>
</tr>
</tbody>
</table>
sits and sits and sits out in the space, it never moves from there, but when we talk about the light, you surly won’t say rare.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>The sun is like a bubbling bristling bowl of hot water.</td>
<td>✔</td>
<td>A</td>
<td>A</td>
<td>touch-temp</td>
<td>even</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The sun is like a white hot fire slowly burning itself out.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The sun is like hot yellow sand that gets wet by waves, goes out, and turns into night.</td>
<td>✔</td>
<td>A</td>
<td>A/R</td>
<td>visual-color/ cycle, cool</td>
<td>L-H</td>
<td>maint</td>
<td>unusual pairing</td>
</tr>
<tr>
<td>The sun is a big, hot cup of hot chocolate with little marshmallow planets.</td>
<td>✔</td>
<td>A</td>
<td>A</td>
<td>touch-temp, (size)</td>
<td>L-H</td>
<td>maint</td>
<td>unusual/ creative elab.</td>
</tr>
<tr>
<td>The sun is a hot yellowish light bulb turning on and off.</td>
<td>✔</td>
<td>A</td>
<td>A</td>
<td>visual-color, light</td>
<td>H-H</td>
<td></td>
<td>uses behav.</td>
</tr>
<tr>
<td>The sun is like the humungous Pacific Ocean swaying back and forth in space.</td>
<td>✔</td>
<td>F</td>
<td>A</td>
<td>visual-size</td>
<td>even</td>
<td></td>
<td>flawed descr. of movement, shape</td>
</tr>
<tr>
<td>The sun is as huge as 95000000 World Trade Centers glued together like an odd rubber ball.</td>
<td>✔</td>
<td>F</td>
<td>A</td>
<td>visual-size</td>
<td>H-H</td>
<td></td>
<td>weak pairing; high salience T</td>
</tr>
<tr>
<td>The sun is like a light bulb making hot yellow heat.</td>
<td>✔</td>
<td>A</td>
<td>A</td>
<td>visual-light, color; touch-temp</td>
<td>H-H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The sun is like a fiery yellow storm of yellowish rays of snow.</td>
<td>✔</td>
<td>A</td>
<td>A/R</td>
<td>visual-color; comp.</td>
<td>L-H</td>
<td>maint</td>
<td>unusual pairing, creative elab.</td>
</tr>
<tr>
<td>The sun is like a yellow fire breathing dragon flying in the sky.</td>
<td>✔</td>
<td>A</td>
<td>A</td>
<td>visual-color; touch-temp</td>
<td>L-H</td>
<td></td>
<td>strong imagery</td>
</tr>
<tr>
<td>The sun is a ball of fire circling around in space.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Qualitative Analysis of Selection of Vehicles

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TALL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>giraffe</td>
<td>3</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dinosaur</td>
<td>1</td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eiffel Tower</td>
<td>1</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empire State Bldg.</td>
<td>2</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>skyscraper</td>
<td>2</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>monument</td>
<td>1</td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sears Tower</td>
<td>1</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>stick</td>
<td>1</td>
<td>L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tree</td>
<td>3</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fleshlump Eater</td>
<td>1</td>
<td>(SK)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>giant</td>
<td>3</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mount Everest</td>
<td>1</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mountain</td>
<td>1</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>world</td>
<td>1</td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sky</td>
<td>3</td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Godzilla</td>
<td>1</td>
<td>SK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>apartment</td>
<td>1</td>
<td>L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>anything could be</td>
<td>1</td>
<td>L</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>STONG</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>iron</td>
<td>1</td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>giant</td>
<td>2</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>metal</td>
<td>2</td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>iron bar</td>
<td>1</td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>elephant</td>
<td>2</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>donkey</td>
<td>1</td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rock</td>
<td>2</td>
<td>L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ex</td>
<td>1</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>superman</td>
<td>2</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wrestler</td>
<td>2</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blood Bottler</td>
<td>1</td>
<td>(SK)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>steel beam</td>
<td>1</td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>muscle</td>
<td>1</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undertaker</td>
<td>1</td>
<td>(SK)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wind</td>
<td>1</td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>troll</td>
<td>1</td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>King Kong</td>
<td>1</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pippi</td>
<td>1</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Longstocking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>QUIET</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mouse</td>
<td>11</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>snake</td>
<td>2</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cat/cat's feet</td>
<td>4</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BFG</td>
<td>1</td>
<td>(SK)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>time freezing</td>
<td>1</td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clouds</td>
<td>1</td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mosquito</td>
<td>1</td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>desert island</td>
<td>1</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clock</td>
<td>1</td>
<td>L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>flower blooming</td>
<td>1</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wind</td>
<td>1</td>
<td>L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>feather</td>
<td>1</td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fog</td>
<td>1</td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nothing</td>
<td>1</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SOFT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cat</td>
<td>2</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>teddy bear</td>
<td>1</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pillow</td>
<td>7</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cotton</td>
<td>3</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>blanket</td>
<td>3</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sophie</td>
<td>1</td>
<td>(SK)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>feather</td>
<td>6</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>velvet</td>
<td>3</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jupiter</td>
<td>1</td>
<td>L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fur</td>
<td>1</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HARD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rock</td>
<td>13</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>concrete</td>
<td>4</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>blue rock</td>
<td>1</td>
<td>(SK)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>castle</td>
<td>1</td>
<td>L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>stone</td>
<td>2</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>diamond</td>
<td>1</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>iron</td>
<td>1</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bricks</td>
<td>2</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>floor</td>
<td>2</td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>volcano</td>
<td>1</td>
<td>L</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>strongman (in the circus)</th>
<th>2</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>chain</td>
<td>1</td>
<td>M</td>
</tr>
<tr>
<td><strong>SMALL</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nail</td>
<td>1</td>
<td>M</td>
</tr>
<tr>
<td>people</td>
<td>1</td>
<td>L</td>
</tr>
<tr>
<td>bug</td>
<td>3</td>
<td>H</td>
</tr>
<tr>
<td>mouse</td>
<td>4</td>
<td>H</td>
</tr>
<tr>
<td>bee</td>
<td>1</td>
<td>H</td>
</tr>
<tr>
<td>molecule</td>
<td>4</td>
<td>H</td>
</tr>
<tr>
<td>atom</td>
<td>3</td>
<td>H</td>
</tr>
<tr>
<td>crumb</td>
<td>3</td>
<td>H</td>
</tr>
<tr>
<td>quarter</td>
<td>2</td>
<td>M</td>
</tr>
<tr>
<td>ant</td>
<td>1</td>
<td>H</td>
</tr>
<tr>
<td>germ</td>
<td>1</td>
<td>H</td>
</tr>
<tr>
<td>pin</td>
<td>1</td>
<td>H</td>
</tr>
<tr>
<td>pea</td>
<td>1</td>
<td>H</td>
</tr>
<tr>
<td><strong>HAPPY</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>clown</td>
<td>6</td>
<td>H</td>
</tr>
<tr>
<td>people</td>
<td>1</td>
<td>L</td>
</tr>
<tr>
<td>sunny day</td>
<td>1</td>
<td>meta/R</td>
</tr>
<tr>
<td>Queen of England</td>
<td>1</td>
<td>L</td>
</tr>
<tr>
<td>king</td>
<td>1</td>
<td>M</td>
</tr>
<tr>
<td>dream</td>
<td>1</td>
<td>meta/M</td>
</tr>
<tr>
<td>smile/smily face</td>
<td>6</td>
<td>H</td>
</tr>
<tr>
<td>baby</td>
<td>3</td>
<td>M</td>
</tr>
<tr>
<td>nightingale</td>
<td>1</td>
<td>M</td>
</tr>
<tr>
<td>bird</td>
<td>2</td>
<td>M</td>
</tr>
<tr>
<td>Mr Happy character</td>
<td>1</td>
<td>(SK)</td>
</tr>
<tr>
<td>magician</td>
<td>1</td>
<td>L</td>
</tr>
<tr>
<td>cat</td>
<td>1</td>
<td>M</td>
</tr>
<tr>
<td>volcano</td>
<td>1</td>
<td>X</td>
</tr>
<tr>
<td><strong>BRIGHT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sun</td>
<td>15</td>
<td>H</td>
</tr>
<tr>
<td>light</td>
<td>3</td>
<td>H</td>
</tr>
<tr>
<td>lamp/light bulb</td>
<td>2</td>
<td>H</td>
</tr>
<tr>
<td>happy face</td>
<td>1</td>
<td>meta/H</td>
</tr>
<tr>
<td><strong>MAD</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bad man</td>
<td>1</td>
<td>X</td>
</tr>
<tr>
<td>Tasmanian Devil</td>
<td></td>
<td>(SK)</td>
</tr>
<tr>
<td>my mom/brother when...</td>
<td>4</td>
<td>H</td>
</tr>
<tr>
<td>devil</td>
<td>3</td>
<td>M</td>
</tr>
<tr>
<td>thunderstorm</td>
<td>1</td>
<td>meta/H</td>
</tr>
<tr>
<td>mean giants</td>
<td>1</td>
<td>M</td>
</tr>
<tr>
<td>sea</td>
<td>1</td>
<td>L</td>
</tr>
<tr>
<td>monster</td>
<td>2</td>
<td>M</td>
</tr>
<tr>
<td>frown</td>
<td>1</td>
<td>H</td>
</tr>
<tr>
<td>wrestler</td>
<td>1</td>
<td>M</td>
</tr>
<tr>
<td>sad face</td>
<td>1</td>
<td>L</td>
</tr>
<tr>
<td>lion</td>
<td>1</td>
<td>L</td>
</tr>
<tr>
<td>Mr Mad character</td>
<td>1</td>
<td>(SK)</td>
</tr>
<tr>
<td>ogre</td>
<td>1</td>
<td>H</td>
</tr>
<tr>
<td>bull</td>
<td>1</td>
<td>H</td>
</tr>
<tr>
<td>cat with a sore tail</td>
<td>1</td>
<td>H</td>
</tr>
<tr>
<td>bear</td>
<td>1</td>
<td>H</td>
</tr>
<tr>
<td>lightning</td>
<td>1</td>
<td>meta/H</td>
</tr>
<tr>
<td>Jack's mom</td>
<td>1</td>
<td>(SK)</td>
</tr>
<tr>
<td>teachers</td>
<td>1</td>
<td>M</td>
</tr>
<tr>
<td><strong>BEAUTIFUL</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>teacher</td>
<td>4</td>
<td>(SK)</td>
</tr>
<tr>
<td>butterfly</td>
<td>4</td>
<td>H</td>
</tr>
<tr>
<td>rainbow</td>
<td>4</td>
<td>H</td>
</tr>
<tr>
<td>Sophie</td>
<td>1</td>
<td>(SK)</td>
</tr>
<tr>
<td>princess</td>
<td>4</td>
<td>H</td>
</tr>
<tr>
<td>flower/rose</td>
<td>7</td>
<td>H</td>
</tr>
<tr>
<td>coral</td>
<td>1</td>
<td>M</td>
</tr>
<tr>
<td>luna moth</td>
<td>1</td>
<td>M</td>
</tr>
<tr>
<td>dress</td>
<td>1</td>
<td>M</td>
</tr>
<tr>
<td><strong>BRIGHT (cont.)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>moon</td>
<td>2</td>
<td>H</td>
</tr>
<tr>
<td>star</td>
<td>1</td>
<td>H</td>
</tr>
<tr>
<td>city lights</td>
<td>1</td>
<td>H</td>
</tr>
<tr>
<td>Sophie's brain</td>
<td>1</td>
<td>(SK)</td>
</tr>
</tbody>
</table>
APPENDIX J

Demographic and Implementation Questions Asked of Treatment Teachers

1. How long have you been teaching?

2. How long have you been teaching second grade students?

3. What training or coursework have you had in gifted education (brief summary or # of hrs)?

4. What, if any, emphasis did you place on instruction in figurative language and/or poetry analysis this year before starting the unit?

5. How did you decide which students to include in the study group?

6. On what dates (approximately) did you administer the pretests and posttests for the study?

7. Please list (by number is fine) the lessons from the unit that you taught.

For the following two items, please write a few sentences in response.

8. LITERATURE WEB -- Please briefly discuss your use of the literature web, including a description of how frequently you used it, how your students responded to it, and your overall assessment of its effectiveness in supporting literary analysis and interpretation.

9. METAPHOR/COMPARISON CHART — Please briefly discuss your use of the metaphor/comparison chart, including a description of how frequently you used it, how your students responded to it, and your overall assessment of its effectiveness in supporting metaphor comprehension.

10. What is your overall assessment of the effectiveness of this unit? What pieces did you find most/least effective, engaging, and challenging for your students? What recommendations would you have for changes to be made to it?
Vita

Catherine Anne Little

Birthdate: May 5, 1973

Birthplace: Fort Belvoir, Virginia

Education:

1997-2001 The College of William and Mary
Williamsburg, Virginia
Ph.D. in Educational Policy, Planning, and Leadership

1996-1997 The College of William and Mary
Williamsburg, Virginia
Master of Arts in Education

1991-1994 The College of William and Mary
Williamsburg, Virginia
Bachelor of Arts

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.