A pilot study of the "Jacob's Ladder Reading Comprehension Program" with gifted and potentially gifted learners in grades 3, 4, and 5

Heather M. French
William & Mary - School of Education

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

Recommended Citation

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.
NOTE TO USERS

This reproduction is the best copy available.

UMI®
A PILOT STUDY OF THE JACOB'S LADDER READING COMPREHENSION PROGRAM
WITH GIFTED AND POTENTIALLY GIFTED LEARNERS IN GRADES 3, 4, AND 5

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
Heather M. French
November 2005
The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.
A PILOT STUDY OF THE JACOB'S LADDER READING COMPREHENSION PROGRAM
WITH GIFTED AND POTENTIALLY GIFTED LEARNERS IN GRADES 3, 4, AND 5

by

Heather M. French

Approved November 2005 by

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

Michael F. Dipaola, Ed.D.

Bruce A. Bracken, Ph.D.
# TABLE OF CONTENTS

Acknowledgements ............................................................................................................v  
List of Tables ....................................................................................................................vi  
Abstract ...........................................................................................................................viii  

Chapter I: The Problem ........................................................................................................1  
  Introduction ............................................................................................................1  
  Statement of the Problem .......................................................................................3  
  Conceptual Framework ..........................................................................................4  
  Purpose and Research Questions ...........................................................................9  
  Synopsis of Methodology .......................................................................................10  
  Contribution to Gifted Education ..........................................................................12  
  Definition of Terms .................................................................................13  
  Human Subjects ...................................................................................................21  

Chapter II: Review of the Related Literature and Research ............................................22  
  Introduction ..........................................................................................................22  
  Critical Thinking ..................................................................................................22  
  Reading Comprehension ......................................................................................41  
  Curriculum Differentiation ...................................................................................50  
  Teacher Effectiveness ...........................................................................................60  
  Conclusion ............................................................................................................72  

Chapter III: Methodology .................................................................................................73  
  Introduction ..........................................................................................................73  
  Participants ...........................................................................................................73  
  Instrumentation ....................................................................................................80  
  Curriculum ...........................................................................................................88  
  Procedures for the Study ......................................................................................93  
  Delimitations and Limitations of the Study ...........................................................101  
  Conclusion ..........................................................................................................104  

Chapter IV: Findings .......................................................................................................105  
  Introduction ..........................................................................................................105  
  Findings Related to Question 1 ..........................................................................105  
  Findings Related to Question 2 ..........................................................................109  
  Findings Related to Question 3 ..........................................................................113  
  Findings Related to Question 4 ..........................................................................117  
  Findings Related to Question 5 ..........................................................................130  
  Findings Related to Question 6 ..........................................................................133  
  Summary of Findings .........................................................................................136  

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Chapter V: Discussion, Conclusions, and Implications ..............................................141
Introduction .....................................................................................................141
The Effects of Using Jacob's Ladder ..............................................................141
The Effects of Organizational Aspects of the Curriculum .........................152
Implications for Research and Practice ...........................................................156
Summary .........................................................................................................161

References ........................................................................................................163

Appendix A: Classroom Observation Scale-Revised.............................................182
Appendix B: Treatment Fidelity Form .................................................................185
Appendix C: Teacher and Student Feedback Forms ...........................................187
Appendix D: Standards Alignment Chart Example .............................................192
Appendix E: Jacob's Ladder Templates for Ladders A-D ....................................195
Appendix F: Mapping of Paul's Elements of Reasoning to Bloom's Taxonomy, Taba's
Concept Development Model, and Creative Problem Solving .......................200
Appendix G: Implementation Guide for Teachers ..............................................202
Appendix H: Emergent Themes and Representative Comments from Teacher
Feedback Form .................................................................................................207
Appendix I: Student Responses to Questions 1-4 on Feedback Form ...............213

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

I would like to express a sincere, heartfelt thank you to the following people for their continued support and encouragement throughout my time at William and Mary:

My committee: Dr. Joyce VanTassel-Baska, Dr. Michael Dipaola, and Dr. Bruce Bracken; the teachers who agreed to participate in my dissertation research; my husband, Kevin; my parents, Clint and Carolyn Massey; my sister, Lauren Jannik; my friends and colleagues: Kim Tyler, Tamra Stambaugh, Bess Worley, and Catherine Little; Dr. John Noell Moore; and my data guru, Annie Feng.
# LIST OF TABLES

1. Research Questions with Corresponding Instrumentation and Data Collection Procedures ..........................................................74

2. Ethnicity of Districts and Study Sample ..........................................................76

3. Identification Matrix for Gifted and Potentially Gifted Students in the First District ..........................................................77

4. Overall Sample by District, School Type, and Grade ..........................77

5. Research Questions, Implementation, Data Collection/Analysis Procedures and Statistical Analysis Alignment ...........................................100

6. Means and Standard Deviations of Student Scores on the TCT ............107

7. ANCOVA Results for Treatment Effects on TCT .....................................108

8. Results of Paired Samples T-test for Pre-/Post-Gains on TCT ...............109

9. Mean and SD Scores for ITBS by Treatment Group and Grade Level ....110

10. ANCOVA Results for Treatment Effect on ITBS ......................................111

11. Results of Paired Samples T-test for Pre-/Post-Gains on ITBS ...............112

12. ANCOVA Results: TCT, Gender x Ethnicity .........................................114

13. Results of Paired Samples T-test for Gender Effects on TCT Student Gains without Classroom 2 ...........................................114

14. ANCOVA Results: ITBS, Gender x Ethnicity .........................................115

15. Results of Paired Samples T-test for Gender Effects on ITBS Student Gains ......115

16. ANCOVA Results: Differential Effects by Grade on TCT and ITBS Student Gains ..........................................................116

17. Descriptive Statistics for Genre, Ladder Type, and Ladder Level ..........118

18. Results of T-tests of Differential Effects by Genre, Ladder, and Ladder Level ...120

19. Emergent Themes from Question 1 on Teacher Feedback Form ............121

20. Results of Teacher Self-Report on Total Number of Readings and Ladders ......122
<table>
<thead>
<tr>
<th></th>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>Teacher Responses to Questions 5-8 on Feedback Form</td>
<td>123</td>
</tr>
<tr>
<td>22</td>
<td>Teacher Rankings of Genres with Corresponding Rationales</td>
<td>124</td>
</tr>
<tr>
<td>23</td>
<td>Teacher Rankings of Ladders with Corresponding Rationales</td>
<td>125</td>
</tr>
<tr>
<td>24</td>
<td>Response to Student Feedback Questions 5 and 6</td>
<td>129</td>
</tr>
<tr>
<td>25</td>
<td>Scores on COS-R and Treatment Fidelity Form for Treatment Group Teachers</td>
<td>131</td>
</tr>
<tr>
<td>26</td>
<td>Paul’s Elements of Reasoning and the Corresponding TCT Items</td>
<td>133</td>
</tr>
<tr>
<td>27</td>
<td>MANOVA Results: TCT pre/post Gains by Treatment Group on Elements of Reasoning</td>
<td>134</td>
</tr>
<tr>
<td>28</td>
<td>Mapping of Thinking Models, Paul’s Elements of Purpose and Point of View, Jacob’s Ladder Ladder and Level</td>
<td>136</td>
</tr>
</tbody>
</table>
A PILOT STUDY OF JACOB'S LADDER READING COMPREHENSION PROGRAM WITH GIFTED AND POTENTIALLY GIFTED LEARNERS IN GRADES 3, 4, AND 5

ABSTRACT

The purposes of this study were to investigate the effectiveness of the Jacob's Ladder Reading Comprehension Program in improving the critical thinking and reading comprehension skills of identified gifted or potentially gifted learners in grades 3, 4, and 5. A quasi-experimental design was used, with a sample of 45 third, fourth, and fifth grade identified potentially gifted students and 34 third and fourth grade identified gifted students in the experimental group as well as 40 fifth grade identified potentially gifted students and 35 third and fourth grade identified gifted students in the comparison group. Both experimental and comparison groups were assessed before and after the eight week intervention on a measure of critical thinking and a measure of reading comprehension. Other data sources included classroom observations, teacher and student feedback, student products, and teacher focus groups.

Findings around critical thinking and reading comprehension did not show statistically significant gains in student learning. However, feedback from teachers and students were overwhelmingly positive, suggesting that further research needs to be conducted to determine if quantitative measures of student learning corroborate this qualitative data. Student performance data in several of the ladders and ladder levels was also statistically significant in a positive direction. Additionally, statistically significant differences were found based on gender. Teacher effectiveness was also negatively correlated with student success with the program. Specific suggestions for future research are provided.
CHAPTER I
The Problem

Introduction

Learning to read and becoming proficient in reading comprehension is a vital component of public education for all students. Comprehension, especially, has been called the "essence of reading" (Durkin, 1993, p. 21) and is considered essential for academic learning as well as lifelong learning. Despite the importance of this skill, analysis of student performance on the National Assessment of Educational Progress (NAEP) shows the majority of fourth and eighth grade students in the United States are scoring at or below the "basic" level of achievement in reading comprehension (NCES, 2004). The "basic" level of achievement is defined as "partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade" (NCES, 2004, p. 2). The interpretation of the NAEP results indicates the majority of the fourth and eighth graders in this country have achieved only partial mastery of fundamental skills for success in academic and lifelong learning. In 2003, only 31% of fourth graders and 32% of eighth graders reached the "proficient" level of achievement with a mere 8% of fourth graders and 3% of eighth graders scoring at the "advanced" level (NCES, 2004).

For students who qualify for free or reduced lunch, the results of the NAEP are even more disheartening. As poverty level increases, the average score on the reading portion of the NAEP decreases (NCES, 2004).

These results are particularly discouraging when a closer examination is made of the types of comprehension skills that are assessed by the NAEP. The majority of
the questions on the NAEP require students to be proficient in only the lowest levels of Bloom’s Taxonomy. They ask students to recall knowledge or information explicitly stated in text, which is level 1 of the taxonomy or to engage skills encompassed in the second level of the taxonomy—comprehension (Bloom, 1956; NCES, 2004). Skills included in the comprehension level of Bloom’s Taxonomy include translating information from one form to another, interpreting text to grasp the theme of the work, the ability to determine the immediate inferences within text based on information that has been explicitly stated, and the ability to make predictions based explicitly on information stated in the text. Some of the questions on the NAEP ask students to engage in skills at the third level of the taxonomy, application, by asking students to make predictions based on hypothetical situations not explicitly stated in the text. However, few or no questions ask whether students engage in skills from the upper levels of Bloom’s Taxonomy: analysis, synthesis, or evaluation. It is these higher level skills that constitute critical thinking, which Paul (1992) states are “increasingly crucial to success” (p. 14). By neglecting these higher levels of the taxonomy and ignoring critical thinking, students are not being assessed on the thinking—and reading—skills that will mostly greatly affect and support their ability to become lifelong learners.

Because critical thinking skills are not directly tested on high stakes state assessments, they also are not being taught in the typical classroom. Students are being shortchanged in their educational experience by not being challenged to think at higher levels. This issue is confounded when considering high ability and gifted students. This subgroup of students is capable of consistently working at the higher
levels of analysis, synthesis, and evaluation. Also, these students often possess a natural curiosity that flourishes when they are given opportunities to think critically (Paul, 1992). Without such opportunities, their enthusiasm for questioning what they read and for asking why may be squelched, which could in turn limit their success as critically thinking adults and contributing members of society (Paul, 1992). More immediately, however, reading comprehension instruction that lacks a critical thinking component denies students, especially those who demonstrate high ability, the opportunity to explore text from multiple perspectives and understand text deeply and more completely.

Statement of the Problem

High ability students in K-12 classrooms are expected to take, pass, and do well on state academic assessments. However, some teachers of high ability or gifted students are realizing that although these students are capable of higher level thinking, they do not always understand the connection between higher level and lower level critical thinking skills (VanTassel-Baska & Bracken, 2005). Because of this potential disconnect, high ability students may not perform as well as expected on state assessments which tend to focus on lower to mid-level thinking skills.

This issue is further complicated in Title I schools where high ability and gifted students often come from economically disadvantaged backgrounds. Because of their limited experience with reading and their overall lower reading achievement scores (NCES, 2004), these students are at a greater risk for not fully developing and exhibiting strong reading comprehension skills; therefore, these students may potentially miss the connection between lower order and higher order critical
thinking. Teachers want to challenge these students by using high level curricula developed specifically for high ability students, but they also know these students must be prepared for the state assessments.

These dilemmas indicate a clear need for a curriculum that bridges lower order thinking and higher order thinking while at the same time preparing high ability and gifted students for state assessments in reading comprehension and challenging them to move higher on the ladder of critical thinking skills. The Jacob’s Ladder Reading Comprehension Program was developed specifically to meet these particular student needs.

Conceptual Framework

Paul’s Reasoning Model

The critical thinking framework on which the curriculum, Jacob’s Ladder Reading Comprehension Program, used in this intervention study is Paul’s (1992) Reasoning Model. Paul (1992) defines critical thinking as

1) Discipline, self-directed thinking which exemplifies the perfections of thinking appropriate to a particular mode or domain of thinking. 2) Thinking that displays mastery of intellectual skills and abilities. 3) The art of thinking about your thinking while you are thinking in order to make your thinking better . . . In critical thinking we use our command of the elements of thinking to adjust our thinking successfully to the logical demands of a type or mode of thinking. (p. 643)
As this definition implies, Paul (1992) places the responsibility for successful critical thinking on the thinker. In order to think critically, students must consciously engage in the metacognitive process of thinking about thinking.

Paul (1992) also examines critical thinking processes within specific disciplines or content areas including reading. He defines a critical reader as someone who “actively looks for assumptions, key concepts and ideas, reasons and justifications, supporting examples, parallel experiences, implications and consequences, and any other structural feature of a written text, to interpret and assess it accurately and fairly” (p. 642). This definition of a critical reader addresses the eight elements of thought that are central to Paul’s (1992) model of critical thinking.

Paul’s (1992) model has three distinct components that collectively engage critical thinking in students: eight Elements of Thought, nine Universal Intellectual Standards, and three Types of Questions. Each component will be considered separately with specific examples that apply to using critical thinking to enhance reading comprehension skills.

The Elements of Thought include: purpose; question at issue; information or evidence; interpretation and inference; concepts and ideas; assumptions of self and of others; implications and consequences; and point of view.

Purpose focuses on the goal of a particular task (Little, 2002; Paul, 1992; Paul & Elder, 2003a, 2003b, 2004a, 2004b). The purpose of a reading comprehension task might be to understand the complexities and nuances of the text and be able to explain the meaning of a reading passage to others.
The question at issue requires students to discern the specific problem or issue of focus within a particular task. An example of question at issue being used during a reading comprehension activity might be to require students to explicitly state the problem being faced by characters in the story, the issue being addressed by the author, or the main idea of a paragraph (Little, 2002; Paul, 1992; Paul & Elder, 2003a, 2003b, 2004a, 2004b).

The information element includes data, facts, observations, and experiences used by students to draw inferences or conclusions about a problem, issue, or solution. For example, while reading and making statements about what they understand to be happening in the text, students should be required to provide justification or factual evidence for these statements. Using information within the text, students should be able to support their answers (Little, 2002; Paul, 1992; Paul & Elder, 2003a, 2003b, 2004a, 2004b).

Interpretations and inferences refer to minor conclusions or judgments made by individuals. Within a reading comprehension task, students are often asked to make inferences about characters’ actions, about what might happen next, and about the author’s intentions (Little, 2002; Paul, 1992; Paul & Elder, 2003a, 2003b, 2004a, 2004b).

The conceptual element in Paul’s (1992) reasoning model includes theories, definitions, models, and/or frameworks used to organize data and give it meaning. The concepts within a reading comprehension task will depend on the text students are reading and might include such ideas as friendship, family, or trust as exhibited through the events and/or character actions in the text. Reading comprehension
questions that ask students to identify the theme of a passage address on the conceptual element of reasoning (Little, 2002; Paul, 1992; Paul & Elder, 2003a, 2003b, 2004a, 2004b).

The element of assumptions refers to the assumptions made by students as they proceed through a task as well as the assumptions of others from whom they gather information. For the purposes of reading comprehension, students must be able to identify the personal assumptions they bring to the task, which are usually based on prior knowledge or experience, as well as the assumptions of the characters within the story and the author who wrote the story (Little, 2002; Paul, 1992; Paul & Elder, 2003a, 2003b, 2004a, 2004b).

Implications and consequences refer to the likely outcomes of a particular course of action. When students are asked to predict what will happen next in a story, they are being asked to make inferences about the results of a specific character’s behavior or actions or the outcome of a particular situation. The determination of implications and consequences also must include potential effects on others not directly involved in the particular scenario under consideration (Little, 2002; Paul, 1992; Paul & Elder, 2003a, 2003b, 2004a, 2004b).

Finally, point of view refers to the ability of an individual to understand his/her own point of view as well as the point of view of others. Similar to assumptions, point of view requires students to examine their beliefs objectively as well as consider how the points of view of others might differ from their own. Reading comprehension tasks that require students to retell the story from another character’s point of view or to make inferences about a character’s feelings are asking
students to separate their own point of view from that of the characters. In order to successfully make this separation in points of view, students must understand their own point of view first (Little, 2002; Paul, 1992; Paul & Elder, 2003a, 2003b, 2004a, 2004b).

All of these elements of thought are influenced by the nine Universal Intellectual Standards of clarity, accuracy, precision, relevance, depth, breadth, logic, significance, and fairness. These intellectual standards should be used to determine quality of reasoning within a given task, including reading comprehension. For instance, clarity requires students to elaborate on statements they make, to give examples, and to use the text to support their assertions. Accuracy focuses on the truth of a particular statement or inference; it is closely linked to the use of data or evidence to support one’s point of view. Precision focuses on the specificity of an answer or statement. Depth asks students to look more deeply at the complexities of a given issue while breadth requires students to examine other perspectives on the same question. Logic focuses globally on the answer as a whole considered alongside the evidence. Significance requires students to determine if they have examined the most important issue, problem, or question. The goal, according to Paul (1992; Paul & Elder, 2003a, 2003b, 2004a, 2004b), is for these intellectual standards to become part of students’ inner voices and infused within their thinking processes.

Paul (1992, Paul & Elder, 2003a, 2003b, 2004a, 2004b) also provides an overview of three kinds of questions, known as one-system, no-system, and conflicting-systems questions, that can be asked of students to engage the nine intellectual standards; these intellectual standards are then applied to the eight
elements of thought thereby promoting active participation by students in the critical thinking process. One-system questions have a definitive answer. They rely on concrete evidence within the text that supports one correct answer. No-system questions require students to make subjective judgments based on their own assumptions, points of view, and inferences. No-system questions are egocentric in the sense that they ask students to only examine their own beliefs, values, assumptions, and points of view, thereby providing a subjective opinion as an answer. Conflicting-systems questions require students to consider evidence from multiple perspectives, which might include conflicting assumptions and points of view. With conflicting systems-questions, there are a range of possible answers from better to worse. Justification of and support for answers become crucial components in critical thinking (Paul, 1992; Paul & Elder, 2003b, 2004a).

Purpose and Research Questions

The purpose of this study was to conduct a pilot study of the *Jacob's Ladder* Reading Comprehension Program, which was developed for high ability learners and was based on Paul’s (1992) Reasoning Model. *Jacob's Ladder* was designed to foster students' progression from lower level thinking skills to higher level thinking skills by encompassing the eight elements of thought. The program has been piloted with identified gifted and identified potentially gifted students in Title I and non-Title I schools.

Specifically, this study will address the following six questions:

1. How much does the use of *Jacob's Ladder* enhance the critical thinking skills of (a) identified gifted learners; (b) identified potentially gifted learners?
2. How much does the use of *Jacob’s Ladder* improve the reading comprehension skills of (a) identified gifted learners; (b) identified potentially gifted learners?

3. How does the effect of the *Jacob’s Ladder* intervention differ by gender, race (white v. non-white), and grade level?

4. How does the effect of *Jacob’s Ladder* differ by genre, ladder type, and ladder level?

5. How is teacher variability related to student performance on *Jacob’s Ladder* tasks?

6. What critical thinking skills are most enhanced by the program?

*Synopsis of Methodology*

*Jacob’s Ladder* was developed with the intention of creating connections between lower level thinking and higher level critical thinking skills through a series of skill “ladders,” each with three “rungs” of questions, which increase in difficulty and abstractness as students “climb” up. Each ladder begins by asking students to answer a factual, text-based, reading comprehension question such as “List, in order, the following events from the story,” or “What details does the author provide about a character in the story.” The second rung of each ladder is more difficult and represents the midpoint between lower level, concrete thinking and higher level, critical thinking focused on one or more of Paul’s (1992) eight elements of thought, which would be the focus of the third and highest rung of each ladder.

The students who participated in the study consisted of experimental and comparison groups of third and fourth grade students who were identified as gifted
and attended a gifted center in a large, suburban district in southeastern Virginia, as well as experimental and control groups of potentially gifted third and fourth grade students who were attending a Title I school in the same district. A comparison group of fifth grade students from a second district was also included. This second district, a large suburban district in northern Virginia, was also participating in the United States Javits Grant, Project Athena. In both districts, students were previously identified as gifted by a committee using a multi-criteria matrix including the following categories: aptitude, achievement, academic performance, a teacher checklist of characteristics of gifted students, and a parent/child checklist. Each component was given a point value based on the student’s scores on the aptitude and achievement tests, his/her grades in Reading or English and math, the total points awarded on the teacher checklist, and the total points from the parent/child checklist. In the first district, any combination of point values that reached the cutoff total of 114 qualified students for the gifted program. None of the components were emphasized more than the other four. The districts do not require the use of any one particular standardized test or checklist. Students who did not meet the requirements but fell less than five percentage points below the cutoff for being identified as gifted were classified as potentially gifted in the first district. In the second district, students are identified as being potentially gifted, or promising learners, when they qualify for school-based services rather than Center-based or pullout programs.

In the first district, students who were identified as gifted were given the option of attending one of four Centers for the Intellectually Gifted that serve gifted learners in self-contained homogeneous classrooms. Students who were identified as
potentially gifted were served in the district's Centers for Enrichment where they received a differentiated curricula and instruction to meet their needs that extend beyond the regular classroom. All the students in the second district were served in school-based models providing differentiated curriculum and instruction in the regular classroom.

Pre- and posttest data was collected on all students' critical thinking and reading comprehension skills. Experimental teachers were asked to note which readings and/or ladders seemed to work well, which seemed to be problematic, and observations about student receptivity and response to the curriculum.

Contribution to Gifted Education

This research builds on a five-year study being conducted by the Center for Gifted Education at the College of William and Mary to measure the effectiveness of a language arts curriculum developed specifically for high ability learners. One of the early findings from this study was the recognition by teachers that many of their students were not quite ready for the high level, critical thinking required by the curriculum. The teachers asked the researchers to provide supplementary material specifically aimed at improving the critical reading skills of their students. Jacob's Ladder was the resulting supplemental material. Since its development, Jacob's Ladder has received many inquiries, requests for training, and initial support within the districts involved in the five-year study as well as at a national conference.

Jacob's Ladder has the potential to provide teachers of high ability students with a curriculum that offers students both practice in reading comprehension skills and challenge for the improvement of their critical thinking skills. However, before it
could be promoted as such, a pilot study to determine the efficacy of this curriculum must be conducted.

*Definition of Terms*

The terms that follow are used often in this study. Some of the definitions apply to the curriculum while others apply to the sample and are indicated as such.

*Gifted Learners*

Giftedness is defined by the U.S. Department of Education (OERI, 1994) as students who “perform or show the potential for performing at remarkably high levels of accomplishment when compared with others of their age, experience, or environment” (p. 26). The state of Virginia, in which this study takes place, defines gifted students as those students “whose abilities and potential for accomplishment are so outstanding that they require special education programs to meet their educational needs” (Virginia Plan for the Gifted, p. 2). The Virginia Plan for the Gifted also recognizes that these students possess talents and abilities that are different from their peers that could potentially remain undeveloped without educational services that differ from what is provided in the regular classroom.

The gifted students in the sample for this study were identified by the first school district using multiple criteria as required by the Virginia Plan for the Gifted. For the development of this plan, the federal and Virginia definitions of giftedness were used. Identification included an evaluation of aptitude (must score at the 85th percentile or above), grade-level achievement tests, classroom performance, teacher recommendation/gifted characteristics, and parental assessment. Students identified as gifted in the first district are served at one of the district's four elementary Centers.
for the Intellectually Gifted which provide full-time programs for students in grades 3-5. Curriculum and instruction in the gifted program are accelerated and enriched in all subject areas.

Potentially Gifted Learners

There is not a federal or state definition of a “potentially gifted” learner; in those districts that identify high ability, promising, or potentially gifted learners, the definitions vary widely. For the purposes of this study, potentially gifted learners will be defined by the criteria used by the districts involved. In the suburban, southeastern Virginia district involved in Project Athena and the piloting of Jacob’s Ladder, potentially gifted learners from diverse populations are served at Centers for Enrichment at the schools for which they are geographically zoned. To be eligible for the enrichment program, students must meet one of the following criteria: 1) be eligible for free or reduced lunch; 2) be from non-traditional families; and/or 3) be a member of a minority culture. These students are identified through recommendations from teachers, administrators, and/or the Gifted Itinerant Teacher and must demonstrate an aptitude score in the 80th percentile or higher. Teacher and parent checklists along with classroom performance at or above grade level and achievement scores in the 80th percentile or higher are also required.

In the large suburban, northern Virginia district involved in Project Athena and the piloting of Jacob’s Ladder, students are identified as promising learners using scores on an aptitude test, a nonverbal intelligence test, achievement tests, and teacher nomination forms. Additional data sources include student portfolios, interviews, and student responses to challenging questions. Students who are identified as potentially
gifted in the second district are served in school-based programs that provide
differentiated curriculum and instruction in the regular classroom.

Low SES Students

Students of low socioeconomic status are defined by school districts as those
students who are eligible for free or reduced lunch. The districts involved in the
current study serves a large population of low SES students. Classification as low
SES for this study will follow the classification of the school districts meaning those
students identified by the district as low SES on the basis of free or reduced lunch
status also will be considered low SES.

White vs. Non-white Race

The majority of the students participating in this study are either Caucasian or
African-American. Less than 5% of the students classify themselves as belonging to
an ethnic group other than these two. Therefore, for the purposes of data analysis and
reporting, all students who are classified as Caucasian will be considered white, and
all students who are classified as African-American or another minority group to
include Asian-American, American-Indian, Hispanic, and multiracial will be
considered non-white.

Reading Comprehension

The National Reading Panel (2000a) defined reading comprehension as the
active process through which “readers derive meaning from text when they engage in
intentional, problem solving thinking processes” (p. 14). The panel elaborates on this
definition by stating, “Reading comprehension is the construction of the meaning of a
written text through a reciprocal interchange of ideas between the reader and the message in a particular text" (National Reading Panel, 2000b, p. 4-5).

While numerous definitions are available for reading comprehension, the National Reading Panel’s (2000a, 2000b) definition of reading as deliberate engagement with the text coincides well with the purpose and process inherent in *Jacob’s Ladder*. When using *Jacob’s Ladder*, students are meant to be actively engaged in the process of “climbing the ladders” from concrete, lower order thinking skills to more abstract, higher order thinking skills. They are continually asked to revisit the text while formulating answers to questions of increasing levels of difficulty. Therefore, the National Reading Panel’s definition accurately describes the process in which students engage while answering comprehension questions about the text selections included in the curriculum.

**Lower Order Thinking/Concrete Skills**

For the purposes of this study, lower order thinking or concrete thinking skills refer to the skills encompassed by the first three levels of Bloom’s Taxonomy: *Knowledge, Comprehension, and Application* (Bloom, 1956), also known as *Remembering, Understanding, and Applying* (Anderson & Krathwohl, 2001). Knowledge or remembering skills involve the recall of specific information, processes, patterns, or structures that have been learned through memorization or that is explicitly stated in the text. Comprehension or understanding skills represent the first level of understanding. Examples of comprehension tasks might include asking students to paraphrase the meaning of a metaphor or simile, to translate information from one form to another, or to summarize a text passage. Application skills—which
begin to build the bridge between lower order and higher order thinking skills—require students to go beyond the text, to begin making predictions about what might happen in hypothetical situations related to the text, or to make generalizations based on knowledge gained from the text (Anderson & Krathwohl, 2001; Bloom, 1956).

Another type of lower order thinking skills operationalized in the Jacob's Ladder curriculum are the strategies developed by Taba (1962). Taba (1962) divides textual content into tiered levels. The first two levels in Taba's (1962) schema coincide with the intended meaning of lower order or concrete thinking skills in this study. The first level is “facts and processes,” which include “ideas at a low level of abstraction, and specific processes and skills” (p. 175). Taba (1962) identifies these facts and processes as fundamental facts that are crucial for progressing to the higher levels of thinking. The second level consists of basic ideas and include ideas about causal relationships, scientific laws, and mathematical principles—the “structure of the subject” (Taba, 1962, p. 176).

While these thinking skills and processes are referred to as lower order and defined as the lower levels of Bloom's Taxonomy and Taba's Levels of Content, they should not be considered less important or unimportant. Concrete thinking skills provide the foundation on which higher order, abstract thinking skills can be built.

Higher Order Thinking/Abstract Skills

Higher order thinking or abstract skills are those skills at the upper three levels of Bloom's Taxonomy, Analysis, Synthesis, and Evaluation, the upper two of Taba’s Levels of Content, and all eight Elements of Thought in Paul’s (1992) Reasoning Model.
The upper three levels of Bloom's Taxonomy include *analysis, synthesis,* and *evaluation*, or, if using the revised Bloom's Taxonomy, *analyzing, evaluating,* and *creating*. Analysis skills include the ability to recognize unstated assumptions, to distinguish facts from hypotheses, to test hypotheses against available data or evidence, and to use the structure or form of text to help gain a better understanding of its meaning. Synthesis or creation skills require students to combine elements or individual parts and create a coherent whole. Demonstration of these skills might include an intact, effective piece of writing, the formulation of an appropriate hypothesis based on a series of data and evidence, and the development of a plan to resolve a given issue. Evaluation, the highest form of thinking in Bloom's Taxonomy and the second highest form of thinking in the revised taxonomy, asks students to make value judgments about information presented in a text. Demonstration of evaluation skills would require students to assess the general reliability and accuracy of a text, to indicate logical fallacies, to compare themes or generalizations, and to compare information against the highest standards in the appropriate arena (Anderson & Krathwohl, 2001; 1996; Bloom, 1956).

The two highest levels of Taba's (1962) Level of Content include *Concepts* and *Thought Systems*. Taba (1962) defines concepts as "complex systems of highly abstract ideas which can be built only by successive experiences in a variety of context" (p. 178). Similar to Bloom’s level of synthesis in that it requires multiple parts, or experiences, to be synthesized into a coherent whole through an understanding of the relationships between the parts. The highest tier of Taba’s (1962) Levels of Content, Thought Systems, includes the methods by which
individuals engage in inquiry. Using definitions, concepts, and principles of a particular field, thought systems direct the "questions asked, the kind of answers sought, and the methods by which they are sought" (Taba, 1962, p. 178). Essentially, a thought system represents a paradigm (Kuhn, 1962). While specific questions regarding thought systems are not included in Jacob's Ladder, students will become familiar with the different methods used by the different genres to convey the message of a given text.

Finally, the eight Elements of Thought in Paul's (1992) Reasoning Model are the higher level thinking skills that guided the development of the Jacob's Ladder Reading Comprehension Program. The Elements of Thought include: the purpose or goal; the question at issue; the point of view or frame of reference; the empirical evidence; the relevant concepts; the assumptions; the inferences; and the consequences and implications that are inherent in a text. These eight Elements of Thought have been discussed in detail in the Conceptual Framework section of this chapter. The connection between the skill ladders of Jacob's Ladder and the eight elements of thought will be discussed in chapter three.

Scaffolding

Scaffolding, a technique encouraged by the design of Jacob's Ladder, is defined as "a strategy which assists learners to extend the current skills and knowledge they bring to the classroom to a higher level of competence" (Kong, 2002, p. 4) or as bridging the gap between where a learner is currently achieving and a higher desirable level of achievement. The concept of scaffolding is based on Vygotsky's (1981) notion of the zone of proximal development which incorporates
his theory of social learning that states mental functioning is not necessarily an individual process but can take place in dyads and groups as well. Children, Vygotsky believed, learn from each other every day (Minick, 1996; Mooney, 2000; Wertsch & Tulviste, 1996).

The zone of proximal development is “the distance between the most difficult task a child can do alone and the most difficult task a child can do with help” (Mooney, 2000, p. 83). Such help, called scaffolding, may come from a teacher, another adult, or from peers who have already mastered the skill (Daniels, 1996). In addition to the zone of proximal development, Vygotsky also addressed the idea of metacognition, which he called inner speech. Inner speech allows individuals to plan and regulate action, and is derived from previous social interaction (Daniels, 1996). The opportunity to learn from peers is inherent in the implementation design of Jacob’s Ladder; teachers can effectively engage students’ zones of proximal development through careful, purposeful grouping of students who can offer to each other a greater understanding of certain concepts or skills. Metacognition is engaged through the student self-assessment process as well as through the internalization of the skill ladders that will occur as students are exposed to multiple readings utilizing the same ladders.

The scaffolding built into Jacob’s Ladder provides foundational, concrete skill ladders that upon which more abstract, higher level skills are added as students “climb” the later ladders. Teachers are also able to target specific skills or skill sets immediately by utilizing the assessment system incorporated into the curriculum. Teachers can provide more or less assistance during completion of the Jacob’s
Ladder tasks as warranted by teacher observation and the student performance data
generated by the assessment forms.

Human Subjects

Approval for conducting research with human subjects is required by the
district where the study will be conducted as well as the College of William and
Mary. If the data indicate a significantly positive effect of the curriculum on
experimental students’ critical thinking and/or reading comprehension skills, control
teachers will be trained on the implementation of Jacob’s Ladder and will be given
the opportunity to use it with their students. All data collected in this study will
remain anonymous and confidential.
CHAPTER II

Introduction

The integral components of the *Jacob's Ladder Reading Comprehension Program* emphasize developing critical thinking skills in 3rd, 4th, and 5th grade students, improving their reading comprehension skills, and providing appropriate curriculum differentiation. The ability of the classroom teacher to effectively deliver the curriculum is also of utmost importance. The review of research and literature for this study will focus on these four broad topics. Initially, critical thinking will be explored broadly and then more narrowly with respect to research and literature focusing specifically on Paul’s (1992) Reasoning Model. Reading comprehension will be explored with respect to its overall importance in education as well as its close relationship to state curricular standards and assessments. Curriculum differentiation will be examined from the perspective of education in general and then more specifically with respect to educating gifted learners. Teacher effectiveness will be addressed in terms of teacher effects on student learning, characteristics of effective teachers, and potential strategies for improving overall teacher quality.

Critical Thinking

Literature

Critical thinking has been defined in many ways by many educators. Stahl and Stahl (1991) identified 28 different definitions of critical thinking with 48 different student abilities that constitute critical thinking. Most of these definitions have several elements in common including: higher level thinking such as analysis, synthesis, and evaluation; an emphasis on deliberateness on the part of the student to
think critically about a text, an idea, or an issue; identification of critical thinking as a type of metacognition, or thinking about thinking; and recognition of the importance of personal assumptions, point of view, evidence, and implications when engaging in the critical thinking process (Dixon, 2002; Little, 2002; Paul, 1992; Stahl & Stahl, 1991; Thompson, 2002).

Regardless of the diverse nature of its definition, critical thinking is considered an essential skill for students to acquire (Dixon, 2002; Gallagher, 1998; Little, 2002; Paul, 1992; Stahl & Stahl, 1991; Thompson, 2002). Little (2002) wrote, “Our attempts to grapple with [life’s] questions, to find direction and support for our actions, are the very processes of reasoning that make us human. Thus, reasoning forms the substance of how we face our environment and learn to live within it” (p. 52). She also argued for the inclusion of critical thinking skills within curricula to add challenge and to provide depth through the exploration of concepts, ideas, and the connections among them. Paul (1992) defines critical thinking as “1) Disciplined, self-directed thinking which exemplifies the perfections of thinking appropriate to a particular mode or domain . . . 3) The art of thinking about your thinking while you are thinking in order to make your thinking better” (p. 643). In other words, regardless of the content, the particular ideas, or the intended outcomes, thinking critically requires a conscious effort on the part of the student. Students must think deliberately and analyze their own thinking while completing the task at hand whether they are solving a problem, analyzing an idea, or reading a text for comprehension. Critical thinking skills are equally important in the core educational subjects.
Perhaps the person most often associated with the concept of critical thinking is Benjamin Bloom. Through the development of his Taxonomy of Educational Objectives, Bloom (1956) brought attention to the varying levels of challenge and abstractness inherent in educational tasks. As Bloom (1956) organized his taxonomy, lower levels represent simple learning behaviors that can be combined together to produce more complex behaviors as educational task demands move through the hierarchy of skills: knowledge, comprehension, application, analysis, synthesis, and evaluation.

Bloom (1956) described the first three levels of the taxonomy as concrete thinking skills while the latter three are described as increasingly more abstract. At any given level within the taxonomy, it is critical to incorporate the skills evident at all levels below it; therefore, when discussing the skill of analysis, Bloom (1956) reiterated the importance of the lower levels of knowledge, comprehension, and application in the analysis process. However, he made a point of differentiating analysis from less complex skills by stating, “Comprehension deals with the content of material, analysis with both content and form . . . One who comprehends the meaning of a communication may not be able to analyze it all effectively” (Bloom, 1956, p. 145). Instead of merely understanding the words and sentences of a text, analysis involves the ability to discriminate fact from hypothesis; to recognize assumptions and points of view; to separate relevant from extraneous information; to identify the relationship between relevant ideas; and to ascertain the author’s purpose (Bloom, 1956).
As students progress in learning through the hierarchy of Bloom’s (1956) Taxonomy, they move from analysis to the fifth level, synthesis. Synthesis is defined as “the putting together of elements and parts so as to form a whole. This is a process of working with elements, parts, etc., and combining them in such a way as to constitute a pattern or structure not clearly there before” (Bloom, 1956, p. 162). This skill differs from the combination of elements that could potentially occur while engaging in skills at lower taxonomic levels because of the creativity required by the synthesis process. Bloom (1956) emphasized that synthesis requires the production of a pattern that was not clearly there before. Although comprehension, application, and analysis might combine elements within a given set of information, it is not until students reach the synthesis level that they are asked to create their own unique product by combining material given in the text with other material such as prior knowledge, personal experience, and theories (Bloom, 1956).

The final and most complex skill in Bloom’s Taxonomy is evaluation. Evaluation tasks students with making value judgments about learning material. Evaluation criteria for “appraising the extent to which particulars are accurate, effective, economical, or satisfying” (Bloom, 1956, p. 185). Bloom (1956) also noted the relationship between the act of evaluating and the affective response of enjoyment, or lack thereof, of an educational task. A difficult task for most people, according to Bloom (1956), is to distinguish between true judgments that have been reached by reasoning and using the skills embedded in the hierarchy and opinions that are reached through a spontaneous reaction to the learning exercise or material. Evaluation is a more complex thinking process than being able to simply express an
opinion. However, Bloom (1956) also noted that evaluation is not necessarily the end of thinking or problem solving; instead, it may be the beginning of another cycle through the hierarchy with new knowledge acquired through the recently concluded thinking process.

In response to growing criticism of the taxonomy, Anderson and Krathwohl (2001) have recently revised Bloom’s Taxonomy in ways they believe better represent the needs of students and teachers in the classroom. The six nouns in the original taxonomy have been replaced with six active verbs to represent thinking as the active process it is. The six revised levels are Remember, Understand, Apply, Analyze, Evaluate and Create. One of the most notable changes in the original taxonomy is the changing of “knowledge” to “remember” to indicate that the lowest level of the taxonomy focuses on listing, describing, recalling, and recognizing information rather than gaining a greater understanding of a concept or idea as the term “knowledge” implies. Another major change is the movement of synthesis, renamed create, to the highest level of the taxonomy and the movement of evaluation, renamed evaluate, to the level just below. Anderson and Krathwohl (2001) consider creating to be the highest level of thinking rather than evaluating because one can judge and critique information without having to be creative or produce something new. However, one cannot be creative or design a new product without first being able to critically judge and evaluate those products that already exist (Anderson & Krathwohl, 2001).

Bloom’s Taxonomy has most often been used by teachers during their planning rather than by students during their learning. However, Little (2002) and
Paul (1992), along with several other scholars (Dixon, 2002; Gallagher, 1998; Paul & Elder, 2003a, 2003b, 2004a; 2004b; Thompson, 2002), believe it is imperative for critical thinking strategies to be directly and explicitly taught to students in a thinking process model. For maximum effectiveness, it is also argued that a particular reasoning or critical thinking model should be adopted and used consistently in order for students to internalize the critical thinking process. While there are several critical thinking models such as Bloom’s Taxonomy (1956) that provide a framework within which teachers can design lessons requiring critical thinking, there are few models that are intended to be used by students as a thinking tool (Little, 2002). One model, developed by Richard Paul (1992) does provide students with a thinking tool which they can apply to multiple content areas including language arts and reading. In order to read well, according to Paul and Elder (2004b), “one must actively construct an interpretation, imagine alternative meanings, imagine possible objections ... one has to assess and judge (criticality) when one reads. Reading is not good reading—accurate clear, plausible—unless it is also critical reading” (p. 40). Using Paul’s (1992) model while teaching reading helps develop these critical thinking skills in students. This model has been used successfully in language arts curriculum for gifted learners (VanTassel-Baska, Zuo, Avery, & Little, 2002), and has been used by The Center for Gifted Education at the College of William and Mary to develop a new reading comprehension curriculum for high ability learners (VanTassel-Baska, French, & Stambaugh, 2004).
Empirical Research: Critical Thinking

Research on the development of critical thinking skills in students is difficult to uncover because of the multitude of interchangeable terms used for these higher level thinking skills. Terms such as reasoning, problem solving, higher order thinking, analysis, and metacognition have been used contemporaneously with the term critical thinking (Burkhalter, 1993; Facione, 1990; Hughes, 2001; Nelson, 1999; Nickerson, 1994). When research is found, it is difficult to complete a comparative analysis since the focus of each study is remarkably different from the others. Finally, research conducted on these concepts is often completed within the field of psychology rather than education, making inferential conclusions about education speculative at best.

Research and development in critical thinking can be organized into three strands: instructional programs for teaching critical thinking, assessing critical thinking, and linking success with critical thinking to the possession of critical thinking dispositions. Each strand will be discussed separately. Then, research specifically focused on the critical thinking model chosen for this study—Paul’s (1992) Reasoning Model—will be explored. Finally, a rational for choosing Paul’s Model will be provided.

Development of Instructional Programs to Teach Critical Thinking

An abundance of instructional programs for teaching critical thinking have been developed in the last several decades (Halpern, 1997; King, 1994). Some of these programs are designed to teach general critical thinking skills while others focus on subject-specific critical thinking skills. The debate over whether general or
subject-specific critical thinking skills are more readily learned or more effectively transferred to other tasks has not been settled. McPeck (1981, 1990) argues that “reasoning skill is not something different from, or over and above, disciplinary thinking,” therefore, “if the disciplines are properly taught, we will get the kind of intelligent thought from students that we normally associate with the phrase critical thinking” (p. 34). McPeck (1981, 1990) firmly believes that critical thinking skills cannot be taught in isolation. Opponents of McPeck argue that there is not enough empirical research to support or deny claims about the generalizability of critical thinking across disciplines and contexts (Norris, 1990); that the act of thinking does not have to be connected to anything in particular any more than the act of “cycling is logically connected to any particular bicycle” (Siegle, 1990, p. 77); and that the everyday issues for which critical thinking skills are crucial are not subject-specific but rather span multiple disciplines and categories of knowledge (Paul, 1990).

Regardless of whether they are meant to teach subject-specific or general critical thinking skills, the vast majority of these critical thinking programs are not based on research. Two programs that have demonstrated effectiveness through empirical evidence include King’s (1989, 1991a, 1991b, 1992, 1994) Guided Reciprocal Peer Questioning and Lipman’s (1988, 1996, 2003) Philosophy for Children Program.

King’s (1989, 1991a, 1991b, 1992, 1994) Guided Reciprocal Peer Questioning began as an instructional design to improve the thinking of college students in response to the increasing demands on institutions of higher education to graduate students who are able to keep pace with the rapidly changing global community (Halpern, 1994). The basic premise of the Guided Reciprocal Peer
Questioning is for professors to provide students with question stems based on the higher levels of Bloom's taxonomy, to model the use of these types of questions, and to eventually ask students to independently create questions for course content that they can then fully answer as a means of studying for assessments. In several studies comparing this instructional method to other popular methods prevalent in college classrooms such as note-taking, writing summaries of material, writing questions without guidance, answering questions provided by the professor, and small group discussion, the Guided Reciprocal Peer Questioning has been found to improve higher level thinking and learning. In addition, a qualitative analysis of the tape recorded discussions among college students participating in these studies revealed a larger focus on inferences, explanations, analysis, synthesis, and evaluation by the Guided Reciprocal Peer Questioning group (King 1989, 1991a, 1991b, 1992, 1994). These students also demonstrated a greater sense of autonomy and increased self-esteem because the students are, according to King (1994), “learning how to approach life in a thoughtful manner” (p. 15).

Lipman’s (1988, 1996, 2003) Philosophy for Children Program uses a series of age-differentiated novels for students with accompanying manuals for teachers that focus on “stimulat[ing] in children patterns of questioning and discussion that are first modeled by the fictional characters in the novels and subsequently continued, by internalization and appropriation by the live children in the classroom, as they talk about what they have learned” (Lipman, 2003, p. 156). The novels target students ages 6-8 through ages 16-17, spanning the majority of elementary school and continuing through high school. The development of the Philosophy for Children
curriculum was based on four underlying presuppositions and seven underlying strategies. The presuppositions focus on the ability of children to focus on the following: abstractions; the acquisition of higher-order cognitive skills as they relate to progression through linguistic stages rather than chronological ages; the importance of experience to cognitive capabilities; and the importance of synthesizing knowledge within and across disciplines. The seven strategies support these presuppositions through improving thinking skills by allowing students to practice and apply these skills; by providing opportunities for students to reflect on their learning; by logically sequencing the curriculum; by providing students with opportunities to respond affectively as well as cognitively; by using novels with characters that are believable representations of real, thinking children; by providing adequate support for teachers; and by providing an adequate number of different perspectives against which students can self-correct their own perspective thereby deepening their comprehension and understanding of information and evidence (Lipman, 2003).

Since the first novel was written in 1969, the Philosophy for Children curriculum has been used by more than 50 countries on all continents (Accorinti, 2000). Several studies have been conducted on Lipman’s (1988, 1996, 2003) Philosophy for Children Program by the New Jersey Department of Education, the Educational Testing Service, and through grants from the Rockefeller Foundation, the National Endowment for the Humanities, Fordham, and the Schultz Foundation. All of the studies have shown some degree of effectiveness by the curriculum for improving the critical thinking, deductive reasoning, and/or higher level thinking
skills of students (Institute for the Advancement of Philosophy for Children, 2004; Naji, 2003). The Institute for the Advancement of Philosophy for Children (2004) also provides a list of 93 research studies that have been conducted on Philosophy for Children in a variety of contexts; all studies have been reviewed by professors at Montclair State University for integrity of design and validity of findings. Nearly all of these studies provide support for the curriculum as an effective means of enhancing students’ critical thinking, reading comprehension, mathematical skills, emotional intelligence, and/or general cognitive ability (Institute for the Advancement of Philosophy for Children, 2004). In 1986, the Philosophy for Children curriculum was designated a “meritorious educational program” by the National Diffusion Network of the U.S. Department of Education; this designation was revalidated in 1995 (Lipman, 1996). In addition, Lipman and his colleagues have created a professional development Institute and a practitioner’s journal to support the use of Philosophy for Children in classrooms.

The success of King’s (1989, 1991a, 1991b, 1992, 1994) and Lipman’s (1988, 1996, 2003) programs illustrate that it is possible to create a curriculum aimed at improving the critical thinking skills of learners.

Assessing Critical Thinking Skills

Another major focus of inquiry with regards to critical thinking is the area of assessment. Attempts to measure critical thinking skills in students, to determine what skills are most relevant, and to develop an instrument that accurately measures the concept it purports to measure have been the focus of several decades of research. In 1993, Ennis completed a review of existing Critical Thinking Assessments and
offered advice to psychometricians who might be interested in creating their own assessment. There are many different instruments that measure critical thinking skills; these instruments fall into three different categories—those that are high stakes critical thinking tests, those that measure several aspects of critical thinking, and those that measure only one aspect of critical thinking. In the first category of tests, high stakes critical thinking assessments, Ennis (1993) includes the *American College Test (ACT)*, the *Medical Colleges Admissions Test (MCAT)*, the *College Board Advanced Placement (AP)* tests, the *Graduate Record Examination (GRE)*, and the *Law School Aptitude Test (LSAT)*. All of these tests are intended for students enrolled in the upper secondary grades or in institutions of higher education. All of these tests have a large research base supporting their validity and reliability; however, Ennis (1993) claims they may not be the best predictors of true critical thinking ability because students’ are often “taught to the test” thereby superficially inflating their scores on these measures.

Some well-known examples of critical thinking tests that measure several aspects of critical thinking include the *Cornell Critical Thinking Test*, the *Ennis-Weir Critical Thinking Essay Test*, the *Ross Test of Higher Cognitive Processes*, and the *Watson-Glaser Critical Thinking Appraisal*. All of these tests were written during or prior to 1985, making their appropriateness for today’s culture somewhat suspect. The various aspects of critical thinking measured by these four tests include understanding the point, identifying reasons and assumptions, stating one’s own point of view, inductive reasoning, deductive reasoning, prediction, explaining fallacies, seeing other possible explanations, avoiding overgeneralizations, completing
analogies, understanding word relationships, interpreting, argument evaluation, and analyzing the attributes of complex stick figures (Ennis, 1993). Only the Watson-Glaser Critical Thinking Appraisal asks students to identify issues or problems and none of the tests ask students to explore multiple points of view or to support claims with evidence (Ennis, 1993; Paul, 1992; Wilson & Wagner, 1981).

Perhaps the most widely used and well-known critical thinking test from the above list is the Watson-Glaser Critical Thinking Appraisal (Watson-Glaser). The Watson-Glaser originated in the 1930s and has undergone over 30 years of research, revision, and development. The major components of the Watson-Glaser measure students' abilities to define a problem, select information that is relevant to the solution, recognized stated and unstated assumptions, develop or select appropriate hypotheses, and to draw conclusions through making inferences (Wilson & Wagner, 1981). Many studies on the reliability and validity of the Watson-Glaser have been conducted. Most have found that, as a whole, the Watson-Glaser has a reliability coefficient of approximately .85 (Landis & Michael, 1981). Content validity as determined by a panel of 12 psychologists found that the majority of items on the Watson-Glaser minimally or somewhat measure the construct of critical thinking (Modieski & Michael, 1983). However, the Watson-Glaser has been criticized for not accurately measuring each subset of critical thinking skills independently of each other. Researchers have been warned about using the Watson-Glaser to measure specific critical thinking skills and encouraged to use it with other measures of critical thinking when making decisions about students on the basis of demonstrated critical

With respect to researchers creating their own measures of critical thinking, Ennis (1993) provides the following advice: If using a multiple choice format, consider incorporating a written justification of the answer to give students to opportunity to demonstrate their understanding of the complexity of engaging in critical thinking; consider creating an essay test with varying degrees of structure from highly structured to minimally structured depending on the purpose for the test (for example, if the score will be used for a high stakes decision, a highly structured essay test would be more desirable; if the purpose is an essay contest, then a minimally structured essay test would be equally valid); and to consider relying on performance assessments rather than using a standardized tests because performance assessments allow for more realistic situations.

Ultimately, Ennis (1993) and other researchers (Cassel & Congleton, 1993; Halpern, 1994, 1997; King, 1994; McPeck, 1981, 1990; Pascarella, 1999) agree that the nebulous and vague definitions of critical thinking make it an extremely difficult construct to measure. For this reason, among others, the critical thinking instrument used in the current study is one that is being used in the larger federal grant, Project Athena, and was created using the same conceptual framework that serves as the foundation for the Jacob’s Ladder curriculum.

Linking Critical Thinking Skills to Critical Thinking Dispositions

While some researchers and theorists focus only on the cognitive skills inherent in the act of thinking critically, others believe gaining the skills is not enough
to make a student a successful critical thinker. Instead, individuals who are successful at critical thinking also possess a set of critical thinking dispositions such as: a “frame of mind” that includes an awareness of the need to evaluate information, a willingness to test assumptions, and a desire to consider all viewpoints (Beyer, 1985); a critical “attitude” that consists of doubt, carefulness, objectivity, and determinism (DeNitto & Strickland, 1987); a desire to use critical thinking skills once they are obtained (Halpern, 1989, 1997); and “traits of mind,” including “independence of thought, fairmindedness, intellectual humility, intellectual courage, intellectual perseverance, intellectual integrity, curiosity, confidence in reason, the willingness to see objections” and the ability to “enter sympathetically into another’s point of view” (Paul & Nosich, 1991, p. 5). Facione (1991) also refers to critical thinking dispositions in his description of the ideal critical thinker:

The ideal critical thinker is habitually inquisitive, well-informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgments, willing to reconsider, clear about issues, orderly in complex matters, diligent in seeking relevant information, reasonable in the selection of criteria, focused in inquiry, and persistent in seeking results . . .” (p. 14)

In other words, according to these researcher and theorists, a person cannot successfully use critical thinking skills without possessing the motivation, perseverance, and desire to actively engage in seeking an ethical, appropriate solution. Passive learners cannot be critical thinkers.
While there are not a large number of empirical studies that specifically investigate the relationship between critical thinking skills and critical thinking dispositions, one such study (Taube, 1995) found support for the use of a two-factor system of critical thinking that includes both skills and dispositions. Taube (1995) administered two measures of critical thinking, three measures of dispositions, and one measure of both critical thinking and dispositions to Purdue University undergraduates while also collecting information about their Scholastic Aptitude Test (SAT) scores and their grade point averages (GPAs). Through a factor analysis, Taube (1995) found evidence to support the notion that “attempts to foster critical thinking address dispositions and attitudes as well as abilities” (p. 27).

Cacioppo and Petty (1982, 1983) found that research participants who demonstrated a high need for cognition, or the tendency to engage in and enjoy thinking, were more likely to excel at tasks requiring critical thinking skills such as discriminating between strong and weak arguments, recalling supporting evidence, and evaluating persuasive arguments. These findings indicate that possessing the disposition of enjoying the thinking process or of desiring cognition is positively linked to one’s ability to successfully use critical thinking skills.

Researchers have also found a tolerance for ambiguity, or the “willingness to accept a state of affairs capable of alternate interpretations,” (MacDonald, 1970, p. 791) to be positively correlated to a proficiency with tasks requiring critical thinking skills; this collection of research suggests that a tolerance for ambiguity is another important critical thinking disposition (Kirton, 1981; Kroll, 1988; MacDonald, 1970; Tegano, 1990).
Empirical Research: Critical Thinking using Paul's Model

Few studies have been conducted that specifically focused on the effectiveness of using Paul's (1992) Model to teach critical thinking skills. Two large-scale studies did use Paul's Model as a focus: Project Phoenix and Project Athena, both out of the Center for Gifted Education at the College of William and Mary.

Project Phoenix encompassed the development, implementation, and analysis of effectiveness of social studies curriculum in grades 2, 4, and 7. One of the research questions addressed whether or not students exposed to "high-powered, interdisciplinary curriculum in social studies" perform better on measures of conceptual and critical thinking than students who are not exposed to this curriculum (VanTassel-Baska, Little, Rogers, Feng, & Drummond, 2002, p. 2). Project Phoenix used an assessment of critical thinking called the CRTA which was developed by an external evaluator and used Paul's model as guide for what critical thinking skills to include. Data from the CRTA analysis over the three-year implementation period found significant gains in critical thinking skills for both gifted and non-gifted students in the treatment group. The Project Phoenix investigators concluded that high level social studies curriculum using Paul's Model as the conceptual basis for critical thinking activities was effective (VanTassel-Baska, Little, Rogers, Feng, and Drummond, 2002).

Project Athena, a five year Javits grant in which the Center for Gifted Education is currently involved, is investigating the effectiveness of language arts curriculum developed for high ability learners in grades 3, 4, and 5. Paul's model is
the basis for several lessons as well as the culminating research project in each unit at
each grade level. Students are asked to reason through an issue identified from a
poem, short story, or novel they have read using all eight elements of thought. Then,
students choose their own issue to research using questions and a research model
targeting the same eight elements. Project Athena uses the Test of Critical Thinking
(TCT) as a pre- and post-assessment tool to measure the critical thinking skills of
students in experimental and comparison groups. The TCT was developed
specifically for the Project Athena research grant, which is also the context for the
_Jacob's Ladder_ study, and is discussed more fully in the Chapter 3. Although data
collection is not yet complete, preliminary results from year 1 of implementation
indicate significant growth gains on the TCT for students in the experimental groups
(VanTassel-Baska & Bracken, 2005).

**Rationale for Using Paul's Model in this Study**

The rationale for using Paul's (1992) Reasoning Model for this study is
threefold: 1) the model includes a comprehensive collection of critical thinking skills
and critical thinking dispositions; 2) the model has been used successfully in
curriculum for gifted learners and is the foundation of the _Jacob's Ladder_ curriculum;
and 3) Paul's (2003b) notion of critical thinking as "self-directed, self-disciplined,
self-monitored, and self-corrective" (p. 1) which makes students responsible for the
quality and consequences of their thinking.

As mentioned, only one test of critical thinking, besides the TCT, asks
students to identify issues or problems (Watson-Glaser) which is an essential
component of Paul's Reasoning Model. In order to thinking critically about an issue,
students must be able to identify the issue first. None of the tests ask students to explore multiple points of view or to support claims with evidence which are also essential elements of the Paul Model. These tests are all based on conceptual models of critical thinking that do not include these essential elements. When engaged with Paul's Reasoning Model, students are asked to approach a problem or issue from multiple perspectives and to analyze potential solutions from these perspectives. Also, when students present possible solutions or argue against alternatives, they must be able to support their position with evidence from the data available to them such as texts, experiment results, or sets of graphs.

In addition to including more critical thinking skills than other conceptions of critical thinking, Paul's Reasoning Model also incorporates important critical thinking dispositions in the form of nine Universal Intellectual Standards—clarity, accuracy, precision, relevance, depth, breadth, logic, significance, and fairness—against which students should measure all their work.

Because the Paul Model incorporates these additional elements of critical thinking and critical thinking dispositions into one coherent model, it provides a more comprehensive approach to helping students develop critical thinking abilities.

Paul's Model has also been used with success in the research-based language arts curriculum developed by the Center for Gifted Education at the College of William and Mary (Center for Gifted Education, 1998a, 1998b, 1998c). Research participants who are currently a part of the Javits federal grant from which the current study emerges have already been exposed to the Paul Reasoning Model. Several of the Centers for the Intellectually Gifted in the southeastern Virginia district in which
this study occurs also use the curriculum from the William and Mary Center for Gifted Education. The success of this curriculum with gifted learners also informed the development of the *Jacob's Ladder Reading Comprehension Program*. In addition, the skill ladders developed for *Jacob's Ladder* are based on Paul’s Model in response to teachers’ requests for materials that more explicitly address the development of critical thinking skills in their students.

Finally, the fact that Paul (1992, 2003a, 2003b, 2004a, 2004b) intends for his model of critical thinking to permeate all aspects of people’s lives speaks to the greater purpose of education and the *Jacob's Ladder* curriculum. The purpose is not only to ensure that students can engage in the use of critical thinking skills or to improve their reading comprehension of the specific texts included in the program; the purpose is to provide students with a thinking model that will inform their reading in the future, their approach to real life issues, their tolerance for other points of view, and their appreciation for well-supported arguments.

Essentially, the three reasons for choosing Paul’s (1992) Model embody an approach to critical thinking that extends beyond the classroom; this approach extends to the everyday lives of students by helping them to become critical thinkers, rather than just teaching them to think critically in isolated situations.

*Reading Comprehension*

*Introduction*

In 1979, Dolores Durkin published her landmark study on classroom reading instruction and the lack of specific instruction in reading comprehension. She found that teachers spent less than 1% of their reading instruction time on comprehension
and, when comprehension was covered, it was merely mentioned rather than explained to or modeled for students (Asselin, 2002; Durkin, 1979). In 1998, Pressley found that despite 20 years of focused attention, students are still not receiving adequate instruction in reading comprehension.

Literature and research in the field of reading instruction has attempted to elucidate the reasons why student achievement in reading has not improved despite years of efforts and what can or should be done to enhance reading instruction.

**Literature**

The literature on reading comprehension, reading comprehension processes, and the research on each has a long history. The majority of the research on the processes involved in reading comprehension has focused on the strategies used by good readers with good comprehension skills and whether or not such strategies can be taught to all students. In their review of the literature, Duke and Pearson (2002) identified approximately 14 processes activated by good readers during the reading process. Good readers are active readers in that they are constantly evaluating, summarizing, questioning, revising their interpretations of the text, and making predictions about what they are reading. They are also discerning; they use different reading strategies for different types of reading. For example, when reading a narrative, good readers will focus primarily on the elements of literature such as the setting and character. In contrast, when reading an expository piece of writing, good readers will engage in continual summarization to refine their interpretations and understandings of the reading material. Good readers are also adept at accessing and incorporating applicable prior knowledge with new information they are gleaning.
from the text. They are able to assimilate and reject knowledge systematically and continuously during the reading process. This assimilation and/or rejection of prior knowledge contribute to the ongoing revision and refinement of meaning which takes place during the process. In other words, good readers are good multi-taskers. They are able to engage in multiple cognitive processes while simultaneously engaged in the decoding of individual words and their meanings (Duke & Pearson, 2002; Snow, Sweet, Alvermann, Kamil, & Strickland, 2002).

However, before these processes can develop in good readers, there are some prerequisite experiences that must occur (Snow et al., 2002). In their report for the Department of Education’s Office of Educational Research and Improvement (OERI), the RAND corporation’s Reading Study Group outlined the following prerequisites for successful reading comprehension: successful initial reading instruction; good oral language skills; well-developed stores of world knowledge; social interactions with literate individuals in the home and the broader community; and rich exposure to literary experiences (Snow et al., 2002). In other words, students who have been read to as young children, have seen the adults in their lives reading, have been given the opportunity for successful reading instruction, and who have been involved in conversations with others, especially adults, since a young age have a better chance of developing effective reading comprehension skills.

The other large body of literature on reading comprehension focuses on effective instructional practices for teaching the reading comprehension skills of good readers. Some of these instructional practices include questioning, predictions, think-alouds, graphic organizers, the formulation of reading goals, student summaries,
focus on text structure, and the use of metacognition (Duke & Pearson, 2002; Pressley, 2002; Snow et al., 2002). Perhaps most relevant to the current study is the literature and research on questioning and graphic organizers. Questioning refers to asking students questions before, during, and after a reading activity as well as varying the types of questions asked. Numerous studies and reviews of the literature have supported the importance of asking multiple types of questions (Anderson & Biddle, 1975; Duke & Pearson, 2002; Levin & Pressley, 1981). As Duke and Pearson (2002) state in the synthesis of research and literature reviews, “the overall findings [are] that students’ understanding and recall [of the text] can be readily shaped by the types of questions to which they become accustomed” (p. 222). Therefore, if the majority of the questions asked during a reading task are basic knowledge recall questions, then students will be more likely to remember these aspects of the text than any other. If, however, students are asked to make inferences about a text, predict the outcome of a story, or summarize an expository piece, they will become more adept at approaching text from multiple perspectives.

Graphic organizers or visual representations of text, such as concept maps, have also been found effective for teaching reading comprehension skills. Duke and Pearson (2002) focus on studies conducted by Armbruster, Anderson, and Ostertag (1987). Armbruster et al. (1987) found the use of several generic visual aids over an extended period of time to be useful in helping students better understand and organize their understanding of content reading in social studies. These studies also found impressive transfer rates from the use of these graphic organizers to new texts and new reading situations (Duke & Pearson, 2002).
Empirical Research

Systematic research on reading first began over fifty years ago (National Reading Panel, 2000). Research topics have ranged from decoding, word recognition, fluency, comprehension, metacognition, and retention (Snow et al., 2002). Other studies have focused more specifically on individual instructional techniques or practices, such as the use of Reciprocal Teaching and Pre-/During/After-Reading Strategies, to improve reading comprehension ability (Lubliner, 2002; National Reading Panel, 2000b; Salch, 1996; Seymour & Osana, 2003).

In 1997, the National Reading Panel was charged with the task of determining the status of research on reading and the effectiveness of a variety of strategies for teaching reading. The Panel initially screened nearly 100,000 studies published on reading since 1966 using pre-established criteria. In order to be fully considered by the Panel, a research study must be published in a refereed journal, be focused on the reading developed in students grades preschool through 12, and use an experimental or quasi-experimental design. In addition, the report of research was required to have sufficient information about the participants, the intervention, the methods, and the outcome measures used (National Reading Panel, 2000a). Based on their initial perusal of the studies and previous work conducted by the National Reading Council, the Panel determined five major areas of research in the field of reading instruction: Alphabets to include phonemic awareness instruction, phonics instruction, and fluency; Comprehension to include vocabulary instruction and text comprehension instruction; Teacher Preparation and Comprehension Strategies Instruction; Teacher
For the purposes of this study, particular attention will be given to the area of comprehension research which includes vocabulary instruction and text comprehension instruction. Through an analysis of 47 studies on vocabulary instruction, the National Reading Panel (2000b) concluded that vocabulary instruction does lead to improvements in reading comprehension; however, a formal meta-analysis was not possible due the wide range of methods and conceptions of vocabulary instruction used in the studies. In general, vocabulary can be learned incidentally in the context of a narrative; repeated exposure to vocabulary words, particularly beyond a single class period; and pre-instruction of vocabulary words prior to reading were all deemed effective by the studies evaluated by the Panel (2000b).

In the area of text comprehension, research supports the conclusion that explicit teaching of reading comprehension strategies “leads to increased learning of the strategies, to specific transfer of learning, to increased retention and understanding of new passages, and, in some cases, to general improvements in comprehension” (National Reading Panel, 2000b, p. 4-6). Eight instructional strategies yielded strong scientific foundations for improvement of reading comprehension: comprehension monitoring, cooperative learning, graphic and semantic organizers, story structure (who, what, when, where, and why), question answering, question generation, summarization, and multiple strategy teaching that combines several strategies for one comprehension task (National Reading Panel,
Specific strategies that have been studied in-depth include Reciprocal Teaching, direct teaching, and metacognition.

Reciprocal Teaching is an instructional method developed by Palinscar and Brown (1984) that groups the teacher with four to six students to read a text together. After reading is complete, the "learning leader" engages the other group members in the four reading strategies of questioning, clarifying, summarizing, and predicting. Working with these four strategies provides students with a type of metacognition that allows them to monitor their own comprehension of a text (Lubliner, 2002; Seymour & Osana, 2003). While Reciprocal Teaching has been found to be an effective strategy for improving reading comprehension, it is imperative that it be implemented properly and understood fully by teachers (Lysynchuk, Pressley, & Vye, 1990; Palinscar & Brown, 1984; Seymour & Osana, 2003). Research has also been conducted on the individual strategies included in the Reciprocal Teaching method. Questioning was explored by Rosenshine, Meister, and Chapman (1996) as a strategy for improving reading comprehension skills and found to be effective independently of the three other Reciprocal Teaching strategies. Lubliner (2002) systematically explored the use of clarifying as an independent strategy with fifth graders. Lubliner (2002) examined three treatments against a control group for each of the following: clarifying, questioning, and a combination of clarifying and questioning. Results of this study indicated significantly improved comprehension of texts and new vocabulary words encountered during reading for the treatment group learning the clarifying strategy alone (Lubliner, 2002). Thus, research to date seems to indicate the four strategies incorporated in the Reciprocal
Teaching method provide effective reading comprehension instruction either individually or in combination.

Direct teaching applies to comprehension strategies that are specifically taught to students for use during certain phases of reading or to meet specific reading goals and/or specific student needs. Pre, during, and after reading strategies have been shown to be effective in improving the reading comprehension skills of intermediate students as well as increasing their interest in reading and their repertoire of vocabulary words. Pre-reading takes place prior to reading a text and includes, for example, prediction strategies and vocabulary strategies. During reading strategies are those strategies used by students as they are reading through a text. These strategies typically involve checking for understanding of main ideas and themes. After reading strategies approach metacognition as their intent is to prompt students to reflect on what they have read and learned (Salch, 1996). Over time, as students become more familiar with the pre, during, and after reading strategies, the more they independently use these strategies while reading. In addition, students stay on task longer during silent reading, show greater interest in texts, and build their vocabulary more readily (Salch, 1996).

Strategies that have been found effective in meeting specific reading goals and/or specific student needs include focused student response, direct instruction, and visual structures (Barton & Sawyer, 2003). Focused student response provides a method for students to share their interpretations of the reading through talking, writing, or drawing. Effective focused student response begins with the teacher modeling the types of questions and writing prompts that will elicit higher level
responses. The main reading purpose met by focused student response is a deepening in comprehension of a text because without a response comprehension invariably remains shallow (Barton & Sawyer, 2003; Guthrie & Anderson, 1999).

Direct instruction of comprehension strategies focuses on students' ability to employ appropriate reasoning strategies, such as making predictions or drawing conclusions" in order to develop a "complex understanding of the ideas in a text" (Barton & Sawyer, 2003, p. 336). The ten direct instruction, comprehension strategies found most effective by Barton and Sawyer (2003) include locating details; comparing and contrasting; summarizing; envisioning character change such as growth; drawing conclusions; determining cause and effect; making predictions; making thematic connections; and taking multiple perspectives.

Visual structures provide students with concrete representations of abstract thinking processes (Alvermann & Boothby, 1986; Barton & Sawyer, 2003) which are particularly helpful when students are asked to organize large amounts of text such as distinguishing important events from trivial details (Barton & Sawyer, 2003).

Studies on the use of metacognition and instruction aimed at improving students' use of metacognition have yielded strong support for the inclusion of metacognitive skills in reading instruction. Students who regularly use metacognitive strategies more often attain higher levels of thinking, are better able to transfer their learning to other tasks, are more capable of self-assessing their comprehension of texts, and are more motivated to learn (Barton & Sawyer, 2003; Chiu, 1998; Kolic-Vehovec & Bajsanski, 2001; Warian, 2003). While strong readers seem to engage in metacognition instinctively, researchers have found that metacognitive strategies can
be taught through moderately intense direct and/or small group instruction (Barton & Sawyer, 2003; Chiu, 1998).

Regardless of the specific strategy or strategies used, general consensus among reading researchers is that comprehension strategies should be directly taught and can be effectively taught to students (Barton & Sawyer, 2003; Brushaber, 2003; Magliano, Trabasso, & Graesser, 1999; National Reading Panel, 2000b, Palinscare & Brown, 1984; Paris, Wasik, & Turner, 1991; Pressley, 1990; Pressley & Afflerbach, 1995; Rosenshine & Meister, 1994).

Curriculum Differentiation

Introduction

Differentiation is quickly becoming one of the most popular buzzwords in education, particularly when the topic is serving students at the extreme ends of the learning continuum in the regular classroom. Curriculum researchers and developers have addressed differentiation from a number of different perspectives, each choosing to focus on a different element. For example, Tomlinson (1999) has focused on the differentiated classroom, maintaining that differentiation is not just something that happens during instructional time but it is pervasive throughout all classroom experiences. Tomlinson (2000) has also addressed the role of instructional leaders in making differentiated classrooms a success through understanding the concept of differentiation; being optimistic about its success; providing optimal conditions for differentiation to occur; providing relevant staff development for teachers; and communicating with parents about what they should expect from their children's classroom experiences. More recently, Tomlinson (2001) has identified what
differentiation is and is not. According to Tomlinson (2001), differentiated instruction is proactive; qualitative rather than quantitative; rooted in assessment; provides multiple approaches to content, process, and product; is student centered; is a blend of whole class, small group, and individual instruction; and is organic meaning it is something that naturally evolves from within the classroom rather than being something that happens to the classroom.

On the other hand, Gregory and Kuzmich (2004) have focused on using quantitative data for student growth and achievement to inform when, how, and why differentiation occurs; their focus is on making differentiation work within the current standards movement.

Nordlund (2003) has focused primarily on using differentiated instruction to meet the needs of students who qualify for special education due to cognitive impairments, attention deficits, learning difficulties, low English proficiency, and being at risk for school failure. Nordlund (2003) firmly believes teachers and other educational staff should view “diversified learning” as a positive experience for all involved, that variety among students in the classroom enhances the “learning climate” for students, and teachers should encourage students to view the entire school as their classroom where learning can take place anywhere at anytime (p. 5).

VanTassel-Baska and Little (2003) discuss differentiation in terms of curriculum; they espouse that differentiated curriculum should be characterized by an accelerated pace, complexity, depth, challenge, and creativity. These characteristics apply to the content of the curriculum with which students are working, the process through which students learn, and the products that result from student learning.
Beyond this multitude of differentiation strategies related to curriculum and instruction, there are many more strategies that focus on particular grouping methods such as homogenous grouping, ability grouping, cluster grouping, flexible grouping, and pull-out programs. All of these strategies are supported by research (Kulik & Kulik, 1992, 1997; Rogers, 1991, 1998; Swiatek & Benbow, 1991; VanTassel-Baska, 1992, 1998).

It is clear from the mass of differentiated strategies highlighted here that two points of clarification must be made. First, an operational definition of curriculum differentiation must be established for the purposes of this study, and secondly, the major differentiation strategies used in the Jacob’s Ladder curriculum must be discussed. Finally, research on the effectiveness and prevalence of curriculum differentiation will be addressed.

Operational Definition

In the gifted education literature, differentiation is defined as “providing enriched and accelerated curricula, classroom experiences that are challenging and open to discussion, opportunities to work with talented peers and project activity with high-level expectations, and a striving for excellence” (Feldhusen & Moon, 1995, p. 105-06). Parke (1995) also includes the important element of including, or subsuming, the regular curriculum within differentiated curriculum for gifted students. Ehlers and Montgomery (1999), drawing from Maker’s (1982) Curriculum Development for the Gifted, emphasize the necessity of “qualitative differences from the general education curriculum in content, process, product, and learning environment” (p. 96). Dinnocenti (1998) uses Renzulli’s (1997) Five Dimensions of
Differentiation to include differentiated instructional strategies, essentially a differentiated teacher, to the list of content, process, product, and classroom environment. She states, "Educators of the gifted and talented have the task of developing and utilizing the five dimensions of differentiation in a consistent and progressive manner to truly address the needs of highly able learners and direct them into choices that challenge their potential" (p. 11).

Many educators in the field of gifted education recognize that differentiation is a necessity for meeting the needs of these unique students. However, they cannot agree on a standard, operational definition of differentiation. In a qualitative study conducted by Tomlinson with middle school teachers, the teachers were quoted as stating, "'Nobody knows what differentiation means'" (Tomlinson, 1995, p. 79). According to Tomlinson (1999), a differentiated classroom is one in which teachers begin where students are, where teachers provide "specific ways for each individual to learn as deeply as possible and as quickly as possible, without assuming one student's road map for learning is identical to anyone else's," where teacher use time flexibly with a range of instructional strategies, and where teachers have a clear understanding of what powerful curriculum and instruction look like (p. 2). It is this definition that best fits the current study.

The design of Jacob's Ladder and the assessment records that accompany each reading selection allow teachers to determine individual students' strengths and weaknesses in each of the targeted skill areas. Teachers can then tailor the assignment of reading selections based on the areas in which students need additional work and/or those areas in which students are most likely to experience success.
Development of student dyads is also facilitated by pairing students with similar strengths or pairing students whose strengths and weaknesses are mirror images of each other.

In addition to Tomlinson's (1999) broad definition of curriculum differentiation, for the purposes of this study it is important to consider differentiation as it applies specifically to reading comprehension instruction. In an article on reading instruction for gifted learners, Kingore (2002) laments the use of multiple choice reading tasks that limit the gifted student’s "opportunities to demonstrate more advanced interpretations" (p. 13). Instead, she recommends using reading comprehension tasks that require students to generate their own responses to questions that promote higher-level thinking (Kingore, 2002). The design of the skill ladders of the Jacob's Ladder Reading Comprehension Program are specifically intended to provide students with opportunities to demonstrate their critical thinking abilities and advanced interpretations of text. With the exception of the lowest level questions of each skill ladder, the questions are completely open-ended and encourage students to use their interpretation and understanding of the text to answer the questions in their own unique manner. The lowest level questions ask for concrete information that is explicitly stated in the text as a means to ensure a strong foundation on which students are building their interpretations.

**Strategies of Differentiation Inherent in Jacob's Ladder**

Three main strategies of differentiation are inherent in the design of the Jacob's Ladder Reading Comprehension Program: Higher level thinking, in-depth
analysis of text, and open-ended task demands (VanTassel-Baska, 2003; VanTassel-Baska & Little, 2003).

Higher level thinking is incorporated in Jacob's Ladder through the series of skill ladders that contain increasingly complex task demands as students move up the ladder (VanTassel-Baska, French, & Stambaugh, 2004). Little (2003) encourages teachers to directly teach and promote the adoption of a strong thinking model that can be used by teachers and students during questioning, analyzing literature, writing and revising, and conducting research. Paul's Reasoning model is the recommended choice as it presents a versatile framework for engaging students in higher level thinking across genres and disciplines (Little, 2003).

In-depth analysis of the text requires students to re-read the text multiple times and to engage in analysis, synthesis, and evaluation at the conceptual level; in-depth analysis should be facilitated by the use of short passages that can be read, analyzed, and re-read during a class period (Little, 2003; VanTassel-Baska, French, & Stambaugh, 2004). The reading selections chosen for the Jacob's Ladder curriculum are short in length and meet several of the criteria for appropriate literature choices for gifted learners as outlined by Little (2003): emphasis on variety, open-ended, demonstrating an ability to inspire contemplativeness, and are intellectually challenging. In addition, the structure of the ladder skill sets requires students to re-read the text each time they move up the ladder of higher order thinking as each question requires students to provide support for their answer from the text (VanTassel-Baska, French, & Stambaugh, 2004).
Finally, open-ended task demands are an important element of successful differentiation because they foster a tolerance for multiple viewpoints and the development of a shared understanding among students (Struck, 2003), particularly within the implementation structure of *Jacob's Ladder* which requires students to work in dyads to discuss their individual responses and then reach consensus (VanTassel-Baska, French, & Stambaugh, 2004). Open-ended task demands also encourage students to further explore the issues being raised by the text or the questions being asked about the text (Struck & Little, 2003).

**Empirical Research**

The research on differentiation practices for gifted learners in the regular classroom is scarce. Studies conducted on the effectiveness of the language arts curriculum developed by the Center for Gifted Education at the College of William and Mary have yielded positive results for the use of performance-based tasks, especially with special populations of learners such as those from economically disadvantaged backgrounds (VanTassel-Baska, Avery, Hughes, & Little, 2000; VanTassel-Baska & Bracken, 2005). These studies have also illuminated the benefits of challenging high ability students to work at higher levels of thought, to complete open-ended tasks, and to engage in real-world problem solving. When these criteria are met through high level content, students show gains in literary analysis and persuasive writing, skills that require strong critical thinking abilities (VanTassel-Baska, Avery, Hughes, & Little, 2000; VanTassel-Baska, Hughes, Avery, & Little, 1996).
Additionally, recent studies conducted by the National Research Center on the Gifted and Talented have provided the greatest amount of empirical evidence regarding classroom practices specifically designed to meet the needs of gifted learners. In 1992, Archambault, Westberg, Brown, Hallmark, Emmons, and Zhang conducted a national survey of third and fourth grade teachers regarding the instructional and curricular methods they use with their gifted and talented learners. The researchers conducting the Classroom Practices Survey developed the Classroom Practices Questionnaire and distributed it to a total of 7400 third and fourth grade teachers working in public and private schools. The sample was stratified to accurately represent the proportion of teachers in public versus private schools as well as teachers in districts with a higher prevalence of minority students. The questionnaire contained 39 close-ended questions covering six different factors: Questioning and Thinking; Providing Challenges and Choices; Reading and Written Assignments; Curricular Modifications; Enrichment Centers; and Seatwork (Archambault et al., 1992; Westberg, Archambault, & Brown, 1997).

Of all the surveys disseminated, the researchers achieved a remarkable return rate of 53.1%. The major findings of the study revealed that teachers are making only minor modifications in curriculum and instruction to meet the needs of their gifted learners. Of those teachers who did provide differentiation for gifted students, the most common modifications were advanced reading, independent projects, enrichment worksheets, and assigning various types of reports. None of these modifications reach the level of depth, complexity, and challenge recommended for gifted learners. Only a very few teachers compacted the curriculum, provided
opportunities for acceleration, asked for gifted students' input on curricular modifications, or exposed students to higher level thinking skills (Archambault et al., 1992; Westberg, Archambault, & Brown, 1997).

The researchers did find, however, that certain teacher and/or classroom characteristics and experiences were correlated with greater prevalence of modifications for gifted learners. For example, teachers who had received training in gifted education through professional development or university coursework were more likely to differentiate for gifted students. Also, a positive relationship was found between the number of gifted students in a classroom with the amount of modifications afforded to all students. In classrooms with five or more gifted students, teachers provided more modifications in the Challenges and Choices category than teachers who had no gifted learners, 1-2 gifted learners, or 3-4 gifted learners. Thus, as Westberg, Archambault, and Brown (1997) state, “when a substantial number of gifted students were in a classroom, enhanced learning opportunities were made available to the entire class” (p. 31).

Based on the results of this study, several follow up studies have been conducted by the National Center for Research on the Gifted and Talented. The Classroom Practices Observation Study (Westberg, Archambault, Dobyns, & Salvin, 1993) was designed to examine, through classroom observations, the instructional and curricular practices being used with gifted learners in regular elementary classrooms. Forty-six third and fourth grade classrooms were observed for two days each across the four major regions of the United States—South, West, North Central, and Northeast—and in suburban, rural, and urban school districts. The observations were
conducted in all major subject areas. The results of this study indicate that little or no
differentiation in the instructional and curricular practices is occurring for gifted
learners. Across all subject areas, students were heterogeneously grouped for 79% of
instructional time. In addition, the majority of instructional time was spent on passive
activities, with 84% of the activities containing no differentiation for gifted and
talented students (Westberg, Archambault, Dobyns, & Salvin, 1993). In a follow-up
study using the same instruments and the same procedures, Westerberg and Daoust
(2003) found similar results: “The major conclusion draw from the replication study
is that teachers’ differentiation practices in third and fourth grade classrooms have not
changes in the last 10 years” (p. 5).

The results of these studies indicate a clear need to proactively
promote differentiation practices particularly since gifted pull-out programs and
resource rooms are becoming things of the past. Teachers must be provided with
support from administrators, with access to and opportunities to collaborate with the
gifted specialist, and with appropriate, targeted professional development.
Administrators must take the necessary steps to understand the importance and
complexity of differentiated curriculum and instruction for gifted learners
(Archambault et al., 1992; Westberg, Archambault, & Brown, 1997; Westberg &
Daoust, 2003). The role of the gifted specialist should be expanded to include
consultation and collaboration time with regular classroom teachers in order to
promote differentiation practices. The addition of these necessary roles to the gifted
specialist’s job description will necessitate the delegation of other responsibilities to
other school personnel. Gifted specialists should also be provided with opportunities
to model differentiated lessons for regular classroom teachers (Westberg, Archambault, Dobyns, & Salvin, 1993; Westberg & Daoust, 2003). Finally, professional development focused on differentiation and providing for the needs of gifted learners should provide teachers with more choices in materials, resources, and products to meet their students’ interest and needs; will aid teachers in recognizing a variety of learning styles and abilities; will help teachers grow personally and professionally as they reflectively examine their instructional philosophies and practices; and will encourage teachers to raise their expectations for student work (Gubbins, 2002).

*Teacher Effectiveness*

*Introduction*

It seems intuitive that teachers have the power to positively or negatively impact student learning. However, this issue has been the subject of much debate in curriculum reform literature and research. Using student achievement as a means of determining teacher effectiveness has divided educational experts: “there is no topic on which opinion varies so markedly as that of the validity of basing teacher effectiveness on student learning” (Shrinkfield & Stufflebeam, 1995, p. 17). In an often cited article, Frymier (1997) states:

Because every person is accountable for his or her own behavior but not for what other people do, teachers must be held accountable for what they do as teachers but not for what their students do as learners. Students are responsible for their own learning. (p. 233)
However, research on the effects of teacher behavior on student learning does not support this claim. Rather, recent research on teachers' ability to affect student learning has overwhelmingly shown that teachers have tremendous power to impact student learning (Darling-Hammond 2004a, 2004b; Mendro, 1998; Sanders, 2000; Sanders & Rivers, 1996; Schalock, 1998; Stronge & Tucker, 1999; Tucker & Stronge, 2005; Wright, Horn, & Sanders, 1997). The question then turns to what specific characteristics are possessed by effective teachers and how can schools cultivate these desirable behaviors? The following section will discuss the major research findings surrounding the issue of teacher effects on student learning as well as what characteristics constitute an effective teacher. Initiatives to ensure teacher quality that have thus far been successful will also be explored.

*Teacher Impact on Student Learning*

Perhaps the most well-known research on teacher effects on student learning is the work using the Tennessee Value-Added Assessment System (TVAAS) database. Developed by William Sanders, formerly at the University of Tennessee, this database seeks to provide educators in the Tennessee public school system with longitudinal data regarding student progress over time. In a description of value-added assessment, Sanders (2000) assert that students, teachers, and schools should not be assessed on average test scores, but instead on the progress made by individual students from year to year. Studies using the database have yielded astounding results regarding the profound impact teachers can have on student learning.

Sanders and Rivers (1996) found that otherwise comparable students who were placed with highly ineffective mathematics teachers for three years in a row
scored from 52 to 54 percentile points below students who were placed with highly effective mathematics teachers for three years in a row. In addition, teacher effectiveness demonstrates an additive and cumulative effect; in other words, while an effective teacher receiving a student from an ineffective teacher’s classroom can help that student make strong academic gains, he/she cannot completely overcome the negative impact of the previous year’s ineffective teacher. Equally interesting are the findings that suggest lower ability students are the first to benefit from effective teaching while higher ability students did not show adequate academic progress with any teachers other than those ranked the most effective (Sanders & Rivers, 1996). Finally, Sanders and Rivers (1996) determined that the teacher effects on student learning do not differ significantly based on student ethnicity or socio-economic status.

Studies involving value-added initiatives in other states have reached similar conclusions as those studies involving TVAAS. Mendro (1998) reports that findings from the Dallas Public Schools are nearly identical to those from Tennessee in terms of the profound effects teachers have on student learning. The Dallas studies further indicate that the residual negative impact of ineffective teachers, also found by Sanders and Rivers (1996), may last through as many as three years of instruction even in classes taught by teachers ranked in the top third of effectiveness (Mendro, 1998).

Equally disturbing are findings in Dallas that “lower achieving students are more likely to be put with lower effectiveness teachers” thereby causing the “negative effects of less effective teachers [to be visited] on students who probably need the
most help” (Mendro, 1998, p. 261). Darling-Hammond (2004a, 2004b) also expresses concern over the inequitable distribution of ineffective teachers in classrooms of students needing effective teachers the most. Her research efforts have focused on the high percentage of under-qualified, ineffective teachers in areas characterized by high minority populations and low socio-economic status (Darling-Hammond, 2000a, 200b, 2004a, 2004b). Darling-Hammond (2004a) asserts that the high number of under-qualified teachers leads to high attrition rates which in turn leads to “high replacement costs for teachers who leave early and high educational costs for the undereducation of students who have not had the benefit of trained and experienced teachers” (p. 1951-2).

In Oregon, current initiatives are focused on pre-service and beginning teachers in an effort to ensure teachers have reached an adequate level of effectiveness before they enter the classroom (Schalock, 1998). The preliminary results of this qualitative study of teacher work samples indicate that beginning teachers have a “sense of personal professional responsibility” and the system in Oregon for the teacher preparation is “accomplishing the positive results intended” (Schalock, 1998, p. 283). This system shows high construct and content validity with the educational standards of the state of Oregon; it also seeks to focus prospective teachers on student learning and to provide them with the dispositions, knowledge-base, and perspective necessary to be successful, effective teachers in today’s standards-based schools (Schalock, 1998). According to McConney (1998), the research on Tennessee, Texas, and Oregon all share one common theme: “teachers are effective if and only if they foster student learning” (p. 230).
The findings of several of these value-added studies have also been discussed in relationship to teacher evaluation and professional development decisions. Wright, Horn, and Sanders (1997) summarize the conclusions drawn from the TVAAS studies by stating, “If the ultimate academic goal is to improve the academic growth of student populations, one must conclude that the improvement of student learning begins with the improvement of relatively ineffective teachers . . .” (p. 66). Sanders and Horn (1998) believe the TVAAS data are helping to accomplish this goal by, “providing the data from which individual professional development plans are drawn . . .” (p. 249), with the professional development goals being directly linked to their students’ achievement on the Tennessee Comprehensive Assessment Program (TCAP) (Sanders & Horn, 1998). Subsequently, Sanders and Horn (1998) assert that teacher evaluation should include an assessment of how well a teacher has accomplished the goals “incorporated into [his/her] professional development plan” (p. 249). In Texas, Mendro (1998) recommends the development of targeted, differentiated professional development to increasing the competency of ineffective teachers while allowing more effective teachers freedom to explore new strategies of interest. In Oregon, Schalock (1998) emphasizes the need for faculty members in teacher education programs to deliberately and consciously choose student teaching placements that will provide pre-service teachers with the “experiences and supervision that allow the prospective teacher to meet the [high] expectations of the program” (p. 272). By doing so, these future teachers will have an opportunity to learn about effective teaching as linked to student achievement before they begin teaching solo.
While discussing clear advantages to value-added assessment, all of these researchers also warn against the potential for misinterpreting and/or misusing student achievement data with respect to teacher effectiveness and accountability. Sanders and Horn (1998) and Wright, Horn, and Sanders (1997) warn against evaluating teachers on the basis of student achievement only. Student progress should be one piece of the teacher evaluation process (Sanders & Horn, 1998; Wright, Horn, & Sanders, 1997). Mendro (1998) encourages careful education of administrators and teachers regarding the underlying concepts of value-added assessment and warns against misguided attempts to reduce data to single numbers ranking teachers based on effectiveness. Schalock (1998) reiterates the importance of perspective in viewing student achievement data in conjunction with teacher effectiveness: “Placing the responsibility [of student learning] on the shoulders or [teachers] should be for the purpose of enhancing performance and continuous improvement, not for punishment” (p. 271). This non-punitive approach to using student achievement data to evaluate teachers is supported by others in the field (Darling-Hammond, 2004b; Sanders & Horn, 1998; Stronge & Tucker, 1999; Tucker & Stronge, 2005; Wright, Horn & Sanders, 1997). The use of these data about teacher performance should lead to ongoing, long-term learning experiences for teachers rather than one-time professional development events (Borko, Elliott, & Uchiyama, 2002; Borko, Mayfield, Marion, Flexer, & Cumbo, 1997).

Darling-Hammond (2004b) and Tucker and Stronge (2005) have also explored the issue of teacher evaluation using student achievement data to determine teacher effectiveness. Darling-Hammond (2004b) describes three school systems in
three different states that have improved student learning through redirecting efforts and funding to the recruitment and retention of highly-qualified teachers. In all three cases, Darling-Hammond (2004b) states that “Rather than spending money on an array of special programs to address the problems created by inadequate teaching, the district decided to create a cadre of well-paid and highly qualified teachers to avoid such problems in the first place” (p. 1075). Tucker and Stronge (2005) present the advantages and disadvantages of four teacher evaluation systems in four states that span a continuum from a predominantly qualitative to a predominantly quantitative approach to teacher assessment. Results from preliminary research on all four systems indicate overall improved teacher quality and in two of the four states a measurable increase in student learning as measured by norm-referenced and/or criterion-referenced standardized tests (Tucker & Stronge, 2005). Tucker and Stronge (2005) state, “Clearly, paying closer attention to teaching practices and their effects on student learning has become standard practice in an effort to improve the quality of teaching and learning” (p. 25).

While the focus of literature and research on the effects of teachers on student learning primarily focuses on the teacher, it should be noted that the context in which the teacher is working is also important. The need for administrative support of improved teaching practices is crucial (Darling-Hammond, 2004a, 2004b; Dipaola & Stronge, 2001; Mendro, 1998; Stronge & Tucker, 1999; Tucker & Stronge, 2005). Dipaola and Stronge (2001) make a strong case for holding superintendents more accountable for affecting student learning, particularly in relation to assessing and ensuring administrator and teacher knowledge, curriculum planning, and instructional
leadership. In a review of policies in 42 states, Dipaola and Stronge (2001) found that few states require a systematic evaluation of superintendents that is closely aligned with current professional expectations for superintendents as defined by the American Association of School Administrators (AASA). The state of superintendent evaluation is, according to Dipaola and Stronge (2001), unacceptable. Rather, they argue that “it is imperative that superintendents be evaluated in a manner that meets all the criteria of good personnel evaluation” which will entail “greater compatibility among evaluation instruments, actual duties of the superintendent, and the standards that guide the profession” (Dipaola & Stronge, 2001, p. 109). The role of building level administrators is also important. In his research in the Dallas public schools, Mendro (1998) found that the “quickest way to change the effectiveness of a school, for better or worse, is to change the principal” (p. 263-4). Principals in effective schools, according to Mendro (1998), are intolerant of ineffective teaching; they expect ineffective teachers to either change or leave.

Qualities of Effective Teachers

Research has clearly shown that effective teachers positively impact student learning while ineffective teachers negatively impact student learning (McConney, 1998; Mendro, 1998; Sanders & Horn, 1998; Sanders & Rivers, 1996; Schalock, 1998; Stronge & Tucker, 1999; Tucker & Stronge, 2005; Wright, Sanders, & Horn, 1997). What characteristics, though, constitute an effective teacher?

Much research has been done in an attempt to determine what qualities and characteristics make an effective teacher (Brophy, 1996; Chan, 2001; Creemers, 1994; Cotton, 1995; Darling-Hammond, 2000a, 2000b, 2004a, 2004b; Ford &
Trotman, 2001; Heath, 1997; Hoy & Spero, 2005; Joffe, 2001; Marzano, 2003; Minor, Onwuegbuzie, Witcher, & James, 2002; Stronge, 2002; Tucker & Stronge, 2005). In What Works in Schools: Translating Research into Action, Marzano (2003) synthesizes his own research and the research of Brophy (1996), Creemers (1994), and Cotton (1995) into three teacher-level factors into which characteristics of effectiveness can be categorized: instructional strategies, classroom management, and classroom curriculum design. According to Marzano (2003), teachers must exhibit effective characteristics in all three categories in order to effectively and positively impact student learning overall: “Effective teachers employ effective instructional strategies, classroom management techniques, and classroom curricular design in a fluent, seamless fashion” (p. 77). In terms of instructional strategies, Marzano’s (2003) position is that more is better. The more instructional strategies a teacher has at his/her disposal, the more flexible he/she can be to meet student needs, the more positively he/she can impact student learning. He also recommends all teachers be provided with and use a school-specific unit framework that employs research-based instructional strategies (Marzano, 2003). With respect to classroom management, Marzano (2003) focuses on the ability of teachers to establish and enforce rules and procedures, impose discipline, foster positive student/teacher relationships, and maintain an emotionally objective mindset when dealing with management issues. The effectiveness of teaching is greatly improved when a teacher is able to successfully employ these four management skills. Finally, Marzano (2003) argues that effective teachers, in addition to strong management and
instructional skills, must also be able to logically organize and clearly communicate content. In order to do so, Marzano (2003) states that teachers must take steps to identify and articulate the specifics of content, to ensure that students have multiple exposures to content, to identify procedures to be mastered, to structure content and tasks using the principle of sameness, and to engage students in complex tasks that require them to address content in unique ways. (p. 120)

Stronge (2002) and Tucker and Stronge (2005) organize the qualities of effective teachers in a slightly different manner. They focus on the characteristics that are prerequisites for effective teaching, those that involve a teacher’s personality, and those that involve classroom management, organization, preparing for instruction, implementing instruction, and monitoring student progress/potential. Characteristics that these researchers (Stronge, 2002; Tucker & Stronge, 2005) consider as prerequisites for effective teaching include verbal ability, educational coursework, teacher certification, content knowledge, and teaching experience. Specific teacher personality traits that seem to positively impact student learning consist of caring, listening, understanding, knowing students, being fair and respectful, socially interacting with students, promoting enthusiasm, motivating learners, having a positive attitude toward teaching, engaging in reflective practice (Stronge, 2002; Tucker & Stronge, 2005). Important characteristics related to organizing and implementing instruction to optimal effect include recognizing the importance of instruction, maximizing instructional time, clearly communicating high expectations of students, sufficiently planning for instructional time, using a range of instructional
strategies, understanding the complexities of teaching, using a variety of questioning techniques, and engaging students in the learning process (Stronge, 2002; Tucker & Stronge, 2005). Finally, Stronge (2002) and Tucker and Stronge (2005) identify the following characteristics related to monitoring student progress and/or recognizing student potential as vital to effective teaching: assigning, reviewing, and assessing meaningful homework; providing ongoing feedback to students through formal, informal, formative, and summative assessments; and recognizing and responding to individual and group differences in learning styles and abilities.

Hoy and Spero (2005) have focused their research regarding teacher effectiveness on pre-service teachers by investigating the development and maintenance of a strong sense of self-efficacy in novice teachers. Citing the work of Bandura (1997), Hoy and Spero (2005) apply the following quotation to the claim that a teacher's self-efficacy can either positively or negatively impact the learning of his/her students: ""The self-assurance with which people approach and manage difficult tasks determines whether they make good or poor use of their capabilities. Insidious self-doubts can easily overrule the best of skills"" (p. 344). Hoy and Spero (2005) found that the self-efficacy of pre-service teachers, or their beliefs about their abilities to positively influence student learning, are highest during student teaching. However, during their first year of teaching, these same teachers' self-efficacy drops, indicating a need for more focused mentoring and support of new teachers to maintain the level of belief-in-self experienced by these teachers during their internships (Hoy & Spero, 2005).
Finally, there is also a group of researchers who have focused on the unique characteristics common to teachers who are effective with gifted students (Chan, 2001; Feldhusen, 1985; Ford & Trotman, 2001; Heath, 1997; Joffe, 2001). Feldhusen (1985) observed that the teachers who were most effective with gifted learners are those that share certain characteristics with the students they teach, namely: high intelligence; achievement oriented; knowledgeable; flexible; demonstrating cultural and intellectual interests; respecting differences; and relating well to other gifted people. In addition to these characteristics, other researchers have identified the following characteristics of effective teachers of the gifted: being imaginative, flexible, and open to change (Chan, 2001; Heath, 1997); being innovative (Chan, 2001; Heath, 1997); demonstrating maturity and self-confidence (Chan, 2001; Heath, 1997); portraying enthusiasm (Chan, 2001; Heath, 1997); an ability to teach higher-level thinking skills and problem solving (Chan, 2001; Ford & Trotman, 2001); developing or selecting methods and materials specifically for gifted learners (Chan, 2001; Ford & Trotman, 2001; Joffe, 2001); exhibiting knowledge of the nature and needs of gifted students (Chan, 2001; Ford & Trotman, 2001; Joffe, 2001); promoting student independence and self-concept (Ford & Trotman, 2001; Heath, 1997); being well-versed in a variety of questioning strategies (Chan, 2001; Ford & Trotman, 2001); exhibiting strong motivation and drive (Heath, 1997); ability to identify gifted learners (Ford & Trotman, 2001); developing an environment where gifted students feel both safe and challenged (Ford & Trotman, 2001); and holding state certification/endorsement in gifted education (Heath, 1997). Additionally and finally, in their research focused on gifted minority students, Ford and Trotman (2001)
modified common characteristics of effective teachers of the gifted in an effort to tailor these successful skills to working with nontraditional gifted learners. These modifications focus on teachers being culturally responsive and include: an awareness and appreciation for cultural diversity; using multicultural resources; fostering an appreciation of students’ cultural differences; and creating an environment where students feel safe to explore and express their diversity (Ford & Trotman, 2001).

Conclusion

Literature and research on critical thinking, reading comprehension, curriculum differentiation, and teacher effectiveness indicate the importance of all four in successful education for all learners, and, specifically, gifted students. The *Jacob’s Ladder Reading Comprehension Program* requires students to complete tasks designed to simultaneously hone their critical thinking skills and enhance their reading comprehension ability. Through the types of reading selections included and the open-ended task demands inherent in the skill ladder sets, *Jacob’s Ladder* embeds appropriate curricular differentiation for gifted learners. Additional differentiation by teachers is encouraged through careful analysis of student data delineated by genre and skill set and through effective, flexible instructional strategies. Finally, *Jacob’s Ladder* has been developed with the goal of instilling in students a strong thinking model relevant to multiple disciplines and real life situations as well as the habits of mind conducive to critical reading.
CHAPTER III

Methodology

Introduction

The purpose of this study was to conduct a pilot of the *Jacob's Ladder* Reading Comprehension Program developed by the Center for Gifted Education at the College of William and Mary as part of a United States Department of Education Javits Grant. This curriculum was developed with the intention of improving high ability students’ reading comprehension skills and enhancing their critical thinking skills through evidence-based strategies in both areas. The following chapter will discuss the participants, the instrumentation, the curriculum, and the procedures used in the pilot study of *Jacob's Ladder*. Table 1 begins with a brief synopsis of each research question including the corresponding instrumentation and data collection procedures used.

Participants

Site Selection

The first district in which this study was conducted was already involved in the larger scale Javits grant, Project Athena, from which the development of *Jacob's Ladder* emerged. The district is a large, suburban district in southeastern Virginia educating approximately 33,200 students at four early childhood centers, 28 elementary schools, nine middle schools, and five high schools. Of the 46 schools in the district, 26 are fully accredited by the state. The demographics of the district in terms of ethnicity are presented in Table 2.
### Table 1: Research Questions with Corresponding Instrumentation and Data Collection Procedures

<table>
<thead>
<tr>
<th>Question</th>
<th>Instrumentation</th>
<th>Data Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How much does the use of <em>Jacob’s Ladder</em> enhance the critical thinking skills of (a) identified gifted learners; (b) identified potentially gifted learners?</td>
<td>• Test of Critical Thinking (TCT)</td>
<td>Pre- and post-test scores for all students in both experimental &amp; comparison groups in grades 3, 4, &amp; 5</td>
</tr>
<tr>
<td>2. How much does the use of <em>Jacob’s Ladder</em> improve the reading comprehension skills of (a) identified gifted learners; (b) identified potentially gifted learners?</td>
<td>• Iowa Test of Basic Skills (ITBS) Survey Battery in Reading</td>
<td>Pre- and post-test scores for all students in both experimental &amp; comparison groups in grades 3, 4, &amp; 5</td>
</tr>
<tr>
<td>3. How does the effect of the <em>Jacob’s Ladder</em> intervention differ by gender, race (white v. non-white), and grade level?</td>
<td>• Test of Critical Thinking (TCT) • Iowa Test of Basic Skills (ITBS) Survey Battery in Reading • Student Demographics • Scored student products</td>
<td>Student pre- and post-test scores separated by gender, race (white v. non-white), and grade level</td>
</tr>
<tr>
<td>4. How does the effect of <em>Jacob’s Ladder</em> differ by genre, ladder type, and ladder level?</td>
<td>• Teacher Feedback Form • Student Feedback Form • Random Selection of Student Products</td>
<td>Teachers and students will complete feedback forms at the end of the implementation (see Appendix C)</td>
</tr>
<tr>
<td>5. How is teacher variability related to student performance on <em>Jacob’s Ladder</em> tasks?</td>
<td>• Classroom Observation Scale-Revised (COS-R) • Treatment Fidelity Form • Test of Critical Thinking (TCT)</td>
<td>A selection of student products will be collected from the beginning and end of implementation; genre and ladder assessment scores will be analyzed. All experimental and comparison teachers will be observed once using the COS-R. Treatment Fidelity form will be used when observing experimental teachers. Pre- and post-test scores of each item of the TCT will be compared; items will be mapped to the critical thinking domain being targeted by the question.</td>
</tr>
<tr>
<td>6. What critical thinking skills are most enhanced by the program?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The average student-to-teacher ratio in the first district is 20-to-1 at the elementary and middle school level and 22-to-1 at the high school level. Approximately 4.6% of the students in this district are identified for the Talented and Gifted (TAG) programs. Of the 33,200 students in the district, an estimated 15,800 are enrolled at the elementary level.

Due to difficulties obtaining a fifth grade comparison sample from the first district involved in this study, a group of fifth graders from another district also involved in Project Athena were included. Choices for a second district were limited to the six other districts involved in Project Athena because posttest scores from the ongoing Athena study were being used as pretest scores for the current study on Jacob's Ladder. The second district was chosen because of its similarities to the first district. Of all the remaining Project Athena districts, this second district was most like the first in terms of size, Title I status, and the large number of students with military dependent status. The comparable of military dependents in the student population was the primary reason for adding the second district to the study. This second district is a large, suburban district in northern Virginia educating approximately 165,000 students at 136 elementary schools, 22 middle schools (grades 6-8), four secondary schools (grades 7-12), and 21 high schools. The demographics of the district in terms of ethnicity are presented in Table 2 to facilitate comparisons to the first district.

The sampling procedures for the study resulted in an experimental group of 75 students and a comparison group of 79 students. The ethnicity most represented in the sample is Caucasian (46.1%), followed by African-American (24%) and Other
(12.3%). Hispanics and Asian-Americans are represented in equal numbers (6.5%), and Native Americans (0.6%) are the least represented ethnic population in the sample. Table 2 presents the demographics of the study sample in terms of ethnicity to facilitate comparisons among the sample and the two districts from which it was drawn. For the purposes of analysis, students were grouped into one of two categories: white or non-white. Students classified as Caucasian were placed in the "White" category and all students classified as African-American, Asian-American, Hispanic, Native American, or Other were placed in the "Non-white" category. These categorizations were made because of the predominance of Caucasian and African-American students in the small sample and the relatively small number of students in the other ethnic categories.

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>First District</th>
<th>Second District</th>
<th>Study Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>African-American</td>
<td>56.9%</td>
<td>10.7%</td>
<td>24.0%</td>
</tr>
<tr>
<td>Caucasian</td>
<td>34.0%</td>
<td>51.4%</td>
<td>46.1%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>5.3%</td>
<td>15.5%</td>
<td>6.5%</td>
</tr>
<tr>
<td>Asian-American</td>
<td>2.6%</td>
<td>17.2%</td>
<td>6.5%</td>
</tr>
<tr>
<td>Native American</td>
<td>0.7%</td>
<td>0.3%</td>
<td>0.6%</td>
</tr>
<tr>
<td>Unspecified/Multiracial</td>
<td>0.5%</td>
<td>4.5%</td>
<td>12.3%</td>
</tr>
</tbody>
</table>

Table 3 presents the sample organized by treatment group, district, school type, grade level, and total number of students.
Table 3: Overall Sample by District, School Type, and Grade Level

<table>
<thead>
<tr>
<th>Treatment Group</th>
<th>District</th>
<th>School Type</th>
<th>Grade Level</th>
<th>Number of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>First</td>
<td>Center for Enrichment</td>
<td>3rd, 4th, and 5th</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Center for Intellectually Gifted</td>
<td>3rd and 4th</td>
<td>34</td>
</tr>
<tr>
<td>Comparison</td>
<td>First</td>
<td>Center for Intellectually Gifted</td>
<td>3rd and 4th</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Second</td>
<td>School-based Program</td>
<td>5th</td>
<td>40</td>
</tr>
</tbody>
</table>

Student Identification as Gifted

Identification of gifted students in the first district entails using a multiple criteria protocol. The protocol includes an aptitude test, an achievement test, classroom performance, teacher recommendation, demonstration of gifted characteristics, and parental assessment. To be identified as gifted, students must score at the 85th percentile or above on either the Otis-Lennon School Ability Test (OLSAT) or the Naglieri Nonverbal Abilities Test (NNAT) in order to continue with the screening process for gifted services. Students must also score at the 85th percentile or higher on the Iowa Test of Basic Skills (ITBS) or the Peabody Individual Achievement Test (PIAT). The final decision regarding identification and placement is made by a committee of the Talented and Gifted (TAG) coordinator for each elementary school, classroom teachers, guidance personnel, reading teachers, and the Itinerant Teacher for Gifted Services (see Table 4). The gifted students in this study were included on the basis of their attendance at the Center for the
Intellectually Gifted chosen by the first district to participate and their enrollment in grades 3 or 4.

Students who are identified as potentially gifted in the first district reside in Title I school concentration areas and must meet one or all of the following criteria: they must be eligible for free or reduced lunch, come from a non-traditional family, or be a member of a minority culture. Other criteria include: scores on the OLSAT or NNAT at or above the 80\textsuperscript{th} percentile; scores on the ITBS or the PIAT at or above the 80\textsuperscript{th} percentile; teacher and parent checklists; and classroom performance at or above grade level in math and/or reading. Students who are identified as potentially gifted are placed in rank order by grade for each of the Centers for Enrichment developed to provide services for potentially gifted, low-income learners. The 20 highest ranking students at each grade level are served at one of the elementary schools providing enrichment services. The potentially gifted students included in this study from the first district were those students in Grades 3, 4, and 5 attending the Title I school already involved in the Project Athena grant. Table 4 presents the identification protocols used to identify gifted and potentially gifted students in the first district involved in this study.
Table 4: Identification Matrix for Gifted and Potentially Gifted Students in the First District

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Gifted</th>
<th>Potentially Gifted</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aptitude Test:</strong> Otis-Lennon School Ability Test (OLSAT) OR Naglieri Nonverbal Ability Test</td>
<td>$\geq 85^{th}$ percentile</td>
<td>$\geq 80^{th}$ percentile</td>
</tr>
<tr>
<td><strong>Achievement Test:</strong> Iowa Test of Basic Skills (ITBS) OR Peabody Individual Achievement Test (PIAT)</td>
<td>$\geq 85^{th}$ percentile</td>
<td>$\geq 80^{th}$ percentile</td>
</tr>
<tr>
<td><strong>Teacher/Administrator Recommendation</strong></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Demonstration of Gifted Characteristics</strong></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Parental Assessment</strong></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Classroom Performance</strong></td>
<td>At or above grade level</td>
<td>At or above grade level</td>
</tr>
<tr>
<td><strong>Meets one of the following criteria:</strong></td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Eligible for free or reduced lunch, from a non-traditional family, and/or from a minority culture</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Identification for gifted services in the second district is based on student scores on the Cognitive Abilities Test (CogAT), the NNAT, the standardized Virginia state assessment, and teacher nominations. Students who demonstrate high abilities are served through one of three modes: differentiated curriculum within the heterogeneous classroom that is developed through collaboration with Gifted and Talented teachers and regular classroom teachers; through pull out programs; or through center-based programs. In addition, a specific model for identifying potentially gifted students from disadvantaged backgrounds has been developed in this district. The aim of this model is to identify students who may have been previously overlooked by adding portfolios, observations, conversations, and student responses to challenging questions to the identification matrix along with the ability and achievement test scores. The students involved in this study all receive school-
based services and are, therefore, all classified as potentially gifted or promising learners.

**Teacher Assignment**

This study included three classes of potentially gifted, low-income students who are currently attending one of the Title I schools participating in Project Athena; these students were randomly assigned to the experimental group prior to the beginning of this federal grant. In addition, two comparison classes from one of the Project Athena school-based programs in the secondary district were involved in this study. Finally, two experimental classes and two comparison classes from one of the district-wide Centers for the Intellectually Gifted in the primary district were added. At this Center, the designation of an experimental or control class was determined by the personnel for the district in collaboration with the researcher.

**Instrumentation**

The instrumentation used in this study was a combination of standardized tests with strong technical adequacy data and instrumentation developed specifically for the study. The two student assessment instruments were used as pretest/posttest measures. The observation form was used to observe both experimental and comparison teachers once during the study with the treatment fidelity form used with experimental teachers only. The teacher and student assessment forms are part of the *Jacob's Ladder* curriculum. Finally, the teacher and student feedback forms were administered and collected at the conclusion of the study.
Test of Critical Thinking (TCT)

The TCT is a test designed by Bracken et al. (2003) at the College of William and Mary to assess critical thinking skills of students in grades 3-5. The manual states, “The TCT presents a balanced framework of critical thinking elements within interesting stories that reflect seven important life-domains for children and adolescents” (p. 1). The TCT is based on the Delphi report (Falcione, 1990) and on Paul’s (1992) Reasoning Model; it utilizes the following operational definition of critical thinking: “the process of making reasoned judgments or inferences about issues or problems based on the evidence available with recognition of the influence of point of view, assumptions, and context” (TCT Examiner’s Manual, p. 5).

The TCT is an easily administered, 45-minute timed-test. Students read 10 different short scenarios and answer a total of 45 multiple choice questions using an answer sheet that can be electronically scanned. The administration of the test is simple, straightforward, and thoroughly explained in the Examiner’s Manual. Sample items are included in the manual with scripted explanations of right and wrong answers. Test administrators are asked to simply read the instructions and samples to students, start the 45-minute testing interval, and stop students from working at the end of the allotted time. Scoring of the TCT produces raw scores for data analysis purposes.

Initial technical adequacy data of the TCT from a pilot study are quite promising. Alpha coefficients for each grade level of the pilot sample are reported as well as the alpha coefficient for the total sample. Internal consistency at Grade 3 is
.85, Grade 4 is .83, Grade 5 is .87; for the total sample it is .89, suggesting a high level of reliability.

Although tests of construct validity have not yet been conducted on the TCT, four other methods of validity have been used—content validity, item content dependence, age/grade progression, and total test ceilings and floors.

For content validity, the TCT was reviewed by individuals knowledgeable about Paul’s (1992) Reasoning Model and the elements of critical thinking. Reviewers also assured that each element of reasoning was uniformly assessed throughout the ten scenarios.

The multiple choice items on the TCT were also reviewed by a panel of individuals who had not had the opportunity to read the scenarios. The purpose of this procedure was to ensure content dependence of the questions. In other words, the developers wanted to make certain students could not answer the questions correctly just by guessing, but would have to carefully read the scenario.

Age/grade progression is an important measure of validity because it can be expected that students will become more adept at critical thinking as they grow older and progress through school. The TCT showed standard deviations at each grade level that are quite consistent, but the pilot also showed students’ mean scores increasing as related to their ages and grade levels, “thus demonstrating that the anticipated age/grade progression further supports the validity of the TCT” (TCT Examiner’s Manual, p. 25).
The final type of validity assessed, and perhaps the most important when considering the use of the instrument with gifted students, is the total test ceilings and floors. The manual states

the test has a strong floor for third-grade high ability students—one that is greater than −2Z. Similarly for high ability fifth-grade students the TCT demonstrates ceilings that exceed +2Z. Given these ceilings and floors, the TCT exhibits sufficient range of difficulty to be an appropriate measure for lower functioning third-grade students and very gifted fifth-grade students.

(p. 26)

Therefore, limited demonstration of student growth on the TCT due to ceiling effects should not be an issue. These measures of validity are sufficient at this stage to justify the use of the TCT with the sample and interventions of this present study.

**Iowa Test of Basic Skills (ITBS) Reading Subtest**

The first ITBS test was administered in 1935 as the Iowa Every Pupil Test; since that time, it has become a mainstay in standardized testing. Intended for use in Kindergarten through eighth grade, the ITBS Core Battery is available in ten levels, 5-14, which roughly correspond to the age of the students taking the test. Each level includes assessments in Vocabulary, Reading, Language, and Mathematics and levels 5-8 have additional assessments in Listening and Word Analysis. The ITBS Complete Battery adds assessments in Social Studies, Science, and Sources of Information beginning with level 7 and provides a Writing and a Listening Assessment for levels 9-14. The ITBS Survey Battery, the option chosen for the current study, is intended to provide a reliable overview of a student’s level of
achievement in Reading, Language, and Mathematics in a minimal amount of testing
time. The Survey Battery accomplishes this task by using a subset of items from the
Core Battery; total administration time for the complete Survey Battery is 100
minutes with 30 minutes allotted for each subtest in reading, language, and
mathematics (Hoover, Dunbar, Frisbie, Oberley, Bray, Naylor, Lewis, Ordman, &
Qualls, 2003).

The "basic skills" measured by the ITBS, per the test authors, are "the entire
range of skills a student needs to progress satisfactorily through school" (Technical
Summary I, p. 15). For the purposes of this study, only the reading portion of the
Survey Battery will be considered. The specific tasks required by this section include
reading passages of various lengths from genres such as fiction, fables, tales, poetry,
interviews, diaries, biographical sketches, science and social studies materials, and
nonfiction other than textbook materials. According to the University of Iowa, the
original creators of the test, approximately 60% of the reading comprehension items
require students to draw inferences or make generalizations about their reading
(University of Iowa, n.d.) The purposes of the ITBS are provided in the Technical
Summary I; the purposes most relevant to this study include aiding teachers in
determining to what extent their students possess the knowledge and skills necessary
for academic success, to estimate the developmental level of students to help teachers
adapt curriculum and instruction to meet individual student needs, to identify student
areas of strengths and weaknesses, to provide achievement information for
estimations of year-to-year progression.
All questions on the ITBS Survey Battery Reading Subtest are in multiple-choice format. Scoring of the ITBS produces raw scores, percent correct scores, developmental standard scores, grade equivalents, national percentile ranks, local percentile ranks, stanines, and normal-curve equivalents. For the purposes of data analysis in this study, the means and standard deviations for the raw scores were used since the researcher was not interested, at this point, in how participants compare with other students in the nation. The raw scores were converted to the IQ metric to facilitate comparisons across grade levels.

The overall technical adequacy of the ITBS is positive. Evidence of reliability was determined using the Kuder-Richardson Formula 20 and includes a reliability coefficient as well as a standard error of measurement for each subtest of the ITBS. In the current study, students in Grade 3 took Level 9, students in Grade 4 took Level 10, and students in Grade 5 took Level 11. Psychometric information has been reported for these three levels only. Reliability coefficients for the ITBS are high with most subtest reliabilities between .85 and .92. During a re-norming of the ITBS during the fall and spring of 2003, the reliability coefficient for the Level 9 reading section ranged from .88 to .89 and the standard error of measurement ranged from 2.2 to 2.1. The reliability coefficient for the Level 10 reading section ranged from .87 to .88 and the standard error of measurement ranged from 2.4 to 2.3. Reliability coefficients for the Level 11 reading section ranged from .86 to .87 with standard error of measurements from 2.5 to 2.4. Correlational validity was determined by comparing scores on the ITBS with scores on tests assessing similar skills such as the
Cognitive Abilities Test (CogAT) and with students’ future grades and test performance (Brookhart, n.d.).

It is important to note the potential ceiling effects when using the ITBS with high ability students. Many gifted students who take the ITBS designated for their chronological age will score perfectly or nearly perfectly because the test is too easy; because of its ease for gifted students, the test has the potential to limit the amount of student growth indicated through pre-test/post-test administration.

**Classroom Observation Scale-Revised (COS-R)**

The Classroom Observation Scale-Revised (COS-R) is an instrument that was designed and revised by VanTassel-Baska, Avery, Drummond, Struck, Feng, and Stambaugh (2004) in an effort to analyze “differences in instructional behaviors seen in different organizational arrangements” (VanTassel-Baska & Feng, 2004, p. 88). Based on research on educational reform, general teaching behaviors, and best practices in classrooms for the gifted, the form is designed to be utilized in all classrooms and in all subject areas.

The COS-R is divided into six subscales. The first subscale focuses on Curriculum Planning and Delivery. Examples of observable characteristics for this subscale include setting high expectations for student performance and asking students to reflect on what they have learned. The second subscale focuses on Accommodations for Individual Differences; examples include accommodating individual differences through materials, conferencing, and/or task assignments and encouraging multiple interpretations. The third subscale focuses on Problem Solving, specifically the heuristic of brainstorming, problem identification and definition, and
developing problem-solutions based on generalizations. The fourth subscale targets Critical Thinking Strategies and includes skills such as engaging students in comparing and contrasting ideas as well as encouraging students to synthesize information within and across disciplines. The fifth subscale focuses on Creative Thinking Strategies; examples include soliciting diverse thoughts about issues or ideas from students and encouraging students to demonstrate open-mindedness and tolerance of imaginative solutions to problems. Finally, the sixth subscale focuses on Research Strategies such as gathering evidence from multiple sources, analyzing data, and encouraging students to identify consequences and implications of their findings. Teacher behaviors are rated as effective, somewhat effective, ineffective or not observed.

In a previous study using the COS-R, overall reliability was .91 and .93 as determined by two separate observations. The reliability of each subscale was also determined and ranged from .67 to .94 (Feng, 2004). Content validity for the COS-R was .98, determined by expert review in an earlier study. The intra-class coefficients for content validity ranged from .85 to .98 (Feng, 2004). A copy of the COS-R can be found in Appendix A.

Jacob’s Ladder Treatment Fidelity Form

Based on the design for implementation of the Jacob’s Ladder intervention, the Treatment Fidelity Form provides the researcher with a list of components that must be included in classroom instruction to ensure treatment fidelity of the reading comprehension program. The Jacob’s Ladder Treatment Fidelity Form was only used when observing experimental classrooms. Each implementation component of
the intervention was rated by the researcher as observed or not observed and comments were added when appropriate. It was expected and acknowledged that only some of these behaviors may be observed by the researcher in a given classroom observation. Feedback was provided to the teacher being observed regarding the strengths and weaknesses of implementation that were evident during the observation. Such feedback occurred face-to-face whenever time and situation permitted but was also provided via email and telephone conversations. A copy of the *Jacob's Ladder* Treatment Fidelity Form can be found in Appendix B.

*Teacher and Student Feedback Forms*

At the conclusion of the curriculum implementation, teachers and students were asked to complete a feedback form. The questions on these forms targeted the issues of receptivity and ease of use. These forms also addressed whether teachers and/or students preferred some genres and ladder types over others and whether some genres or ladder types were more useful in the classroom. Copies of the teacher and student feedback forms appear in Appendix C.

*Curriculum*

*The Jacob's Ladder Reading Comprehension Program*

*Jacob's Ladder* has been developed by the Center for Gifted Education at the College of William and Mary to meet the reading comprehension needs of high ability learners who are participating in Project Athena, a five year curriculum intervention study funded by the U.S. Department of Education, Javit's Program. *Jacob's Ladder* is intended to aid students in moving from basic, concrete comprehension skills to more abstract, higher level, critical reading skills while using
the same text as the basis for questions at each level. Curriculum that deliberately moves students from concrete to abstract thinking enhances reading comprehension skills (Fielding & Pearson, 1994; Villaume & Brabham, 2002) and promotes greater reading growth (Taylor et al., 2003; Knapp et al., 1995). *Jacob's Ladder* was designed to do both—to enhance reading comprehension through a series of “skill ladders” and to promote higher level thinking by using Paul’s (1992) reasoning model as guidance for the types of questions asked in each “ladder” (VanTassel-Baska et al., 2004). In addition, *Jacob’s Ladder* asks students to answer a series of questions based on the same text therefore requiring students to re-read the texts. Re-reading passages leads to improvement in the overall accuracy of students’ comprehension skills (Rawson, Dunlosky, & Thiede, 2000).

The texts for *Jacob’s Ladder* are comprised of 10 passages in each of five genres: nonfiction, myths or fables, poetry, short stories, and essays. The reading selections at each grade level are aligned with national standards in language arts, social studies, mathematics, and science depending on the content emphasis of the text. A standards alignment chart is presented in Appendix D. Each grade level features poetry, nonfiction, and either myths and fables, short stories, or essays for a total of 30 reading selections and accompanying questions. The questions for each selection are organized into one of the four types of “skill ladders” included in the program.

The reading comprehension questions that accompany each reading selection are written in a “ladder” format with three “rungs;” the most concrete question serves as the lowest “rung,” the middle “rung” bridges the concrete and the abstract
questions, and the highest “rung” focuses on higher order critical thinking questions (VanTassel-Baska, French, & Stambaugh, 2004). An example of each of the four ladders can be found in Appendix E.

Skill Ladder A begins with concrete questions focusing on sequencing the events in a particular text, moves to identifying cause and effect, and leads to higher level, critical thinking about the consequences and implications that can be inferred from the passage. This ladder aids students in developing predictions and forecasting skills by requiring them to make connections among data to ascertain what might happen next.

Skill Ladder B begins by asking students to identify details about plot, setting, characters, or other literary elements present in the reading selection. The second rung on this ladder asks students to classify the details they have identified according to similarities and differences. The final, highest rung of Skill Ladder B requires students to make generalizations about the text based on the evidence they have accumulated through answering the questions within the first two rungs; students are engaged in the consideration of details and the relationships among these details in order to arrive at sound conclusions about the text (VanTassel-Baska et al., 2004).

Skill Ladder C begins at the concrete level by asking students to identify the context in which the selected reading occurs or to identify certain qualities of a particular character in a story, fable, myth, or poem. Then, students are asked to make inferences about the context, setting, or characters in the text. They must use evidence from the text to support their inferences and make judgments based on the information provided in the passage. The final rung on Skill Ladder C requires
students to state the central theme or the main idea of their reading, often asking them to determine what the text means.

Skill Ladder D asks students to move from the lower order skill of paraphrasing short passages or quotations from the text. Students are then asked to summarize the reading selection either by identifying the main idea or rewriting the selection including only the most salient information. The final rung of this skill ladder requires students to engage in creative synthesis by creating a unique product based on the reading selection, prior knowledge, and prior reading experience. A creative synthesis question might ask students to write another story, essay, or poem focusing on the same main issue or it might ask students to rewrite a reading selection from another point of view.

Each skill ladder included in Jacob's Ladder provides a structured path for students to follow as they move from concrete thinking about the words, sentences, and literal meaning of text to more abstract thinking about the issues, assumptions, concepts, points of view, and purpose embedded within the same text. This type of organization provides scaffolding to support students as they move from one level of thinking to the next (Gallagher, 1998).

The foundation and format for the ladder questions is based on Paul's Reasoning Model, Bloom's Taxonomy, Taba's Concept Development Model, and/or the Creative Problem Solving Model. The questions on the lowest rungs of the four ladders may focus on one or more of the following: Identifying evidence, purpose, assumptions, point of view, and issue from Paul's Model; remembering, understanding, and applying from Bloom's Taxonomy; detail identification from...
Taba's Model; and mess finding and data finding from creative problem solving. The lowest rung of Ladder A, "Sequencing," asks students to recall the order of events in the text and put them in the correct order, requiring the skill of remembering and understanding from Bloom's Taxonomy. The lowest rung of Ladder B, "Details," asks students to list details from the text or details related to the topic of the text which is the first step in Taba's Concept Development Model.

The middle rungs of the question ladders focus on the following skills: inferences, evidence/data, and point of view from Paul's Model; understanding, applying, and analyzing from Bloom's Taxonomy; classification/categorization from Taba's Model; and data finding, problem finding, and idea finding from creative problem solving. For example, the middle rung of Ladder B, "Classifications," asks students to categorize the details they listed in the first question which is the second step in Taba's Concept Development Model. The second rung on Ladder C asks students to provide evidence or data—one of Paul's eight elements—to support a claim they are making about the text.

The highest rungs of the four ladders target higher order skills from each of the models to include: all eight of Paul's Elements; analyzing, evaluating, and creating from Bloom's Taxonomy; generalizations from Taba's Model; and idea finding and solution finding from creative problem solving. For example, the highest rung of Ladder C, "Theme/Concept," asks students to identify the theme or concept from the text which targets both Paul's Model and Taba's Model. The highest rung on Ladder D, "Creative Synthesis," requires students to use what they have already evaluated from the text to create an original product; these skills are congruent with
Bloom’s Taxonomy levels of evaluating and creating as well as the idea and solution finding steps in creative problem solving.

The mapping chart found in Appendix F represents the alignment between the question ladders and the four critical thinking/creative thinking models used as the foundation for these questions.

*Procedures for the Study*

*Teacher Training*

Prior to using *Jacob’s Ladder* with their students, all experimental teachers were involved in professional development on implementation of this curriculum. Teachers attended a three-hour training during which the conceptual framework, the intended purpose, and the research base of the curriculum were explained. Teachers also had an opportunity to work with the curriculum as a group. Opportunities to ask questions about the curriculum as well as its implementation were provided.

Teachers were then guided through the structure of implementation that was recommended to be used in this study.

*Implementation Guidelines*

For the potentially gifted students who are already participating in the larger federal grant, the ITBS and TCT posttest scores for the grant’s curriculum intervention were used as the pretest scores for the current study. Implementation of *Jacob’s Ladder*, therefore, could not begin until the implementation of the William and Mary Center for Gifted Education curriculum had been completed. Teachers at the Centers for Enrichment began using the curriculum after their training on *Jacob’s Ladder* was complete and approximately 16 weeks after the implementation of the
Project Athena curriculum had ended. The teachers at the Center for Intellectually Gifted began implementation with the administration of the TCT and ITBS pretests. Once the pretests were administered, teachers immediately began using *Jacob's Ladder* with their students.

Teachers were asked to model the implementation process with the first reading selection in the first genre section of the *Jacob's Ladder* notebook. Teachers read the selection, began at the bottom rung of the first corresponding ladder, and worked through the rungs to the highest level. Then, the teacher moved on to the next corresponding ladder, began at the lowest rung, and moved through the questions to the highest rung of the skill ladder. Students were given the opportunity to ask questions regarding the process and the expectations of the reading comprehension program. Teachers then divided their students into pairs or small groups to begin working on the first reading selection in the second section of their *Jacob's Ladder* notebooks. Each time students completed a reading and its corresponding ladders from *Jacob's Ladder*, they answered the questions independently, then discussed their answers with their partner or small group. Independent completion of the *Jacob's Ladder* reading was completed during class time and/or as homework assignments. During the discussion of their individual responses, students worked toward consensus regarding the best answer to each question using all, some, or none of each partner's or group member's individual answers. With some ladder, teachers asked students to share their consensus answers with the entire class and to field questions about their answer choices.
Upon completion of each reading assignment, students were asked to complete the student assessment form corresponding to each reading and each skill ladder associated with it. The teacher was asked to assess each student's answers using the scale provided with the curriculum. On this scale, a score of 0 equals "Needs Improvement," a score of 1 equals "Meets Expectations," and a score of 2 equals "Exceeds Expectations." Teachers regularly reorganized student groups based on student strengths and weaknesses and/or based on student interest of a particular genre or content area. Teachers were also asked to maintain the class assessment sheet that collects all student assessment scores for each reading selection and corresponding ladders on one spreadsheet. However, none of the experimental teachers completed this record sheet.

Teachers worked with students on one to two reading selections and their corresponding ladders each week for a total of eight weeks. The total instructional time allowed for students to complete each reading selection and its corresponding ladders ranged between 30 to 120 minutes, depending on the length of the text and the student discussions. Teachers tailored the assignment of reading selections and ladder sets to best meet the needs of their students, therefore, some classrooms completed more total readings and ladders than other classrooms. The total number of readings completed ranged from seven to 19 with the total number of ladders completed ranging from 14 to 30. The reading selections chosen by teachers and completed by students evenly represented all genres included in the Jacob's Ladder Reading Comprehension Program except for nonfiction. Due to the length of the
nonfiction reading selections, all teachers asked students to complete approximately one to four fewer nonfiction readings than fiction or poetry readings.

The implementation instructions that were provided to teachers can be found in Appendix G.

*Pre-Assessment*

Prior to working with the curriculum for the first time, students were administered the ITBS Reading subtest and the TCT. Both experimental and comparison students completed these tests. Demographic information about each student was also collected to include gender, race, and grade level.

*Classroom Observations*

Each experimental and comparison classroom was observed once during the implementation of *Jacob's Ladder* using the COS-R. The treatment fidelity form was also used when observing experimental classrooms. The observation occurred approximately halfway through the implementation of the curriculum. In addition, several teachers emailed the researcher on a regular basis to provide updates and to ask questions about the implementation process. The classroom observation was intended as a means of ensuring treatment fidelity and to control for the external validity threat of diffusion of treatment from the experimental group to the control group. The data collected during the classroom observations was also used to address the question of teacher variability and its effects on student performance on the *Jacob's Ladder* tasks.
Post-Assessment

Upon the completion of the curriculum, both experimental and comparison groups were administered the TCT and the ITBS Reading subtest again. The time between pretest and posttest was approximately 16 weeks for the identified potentially gifted groups and approximately eight weeks for the identified gifted groups. For both groups, there was at least eight weeks between the pretest and posttest to control for the threat of students becoming “test smart” due to repeated exposure within a short time frame. Teachers and students were also asked to complete a feedback form with targeted questions related to receptivity and ease of implementation.

Data Analysis by Research Question

For each statistical analysis conducted in this study, a $p$ value of .10 was used. This $p$ value was chosen over more conservative values, such as .05 and .01, because the current research is exploratory. In addition, no high stakes decisions will be made about students on the basis of the data collected and analyzed for this study.

Question 1: The TCT pretest and posttest raw scores were analyzed using an Analysis of Covariance (ANCOVA) to control for initial differences between the experimental and comparison groups. The contrast between the pretest and posttest scores on the TCT for each participant subgroup was used to determine the effectiveness of Jacob's Ladder in enhancing students’ critical thinking skills.

Question 2: The ITBS pretest and posttest raw scores were analyzed using an ANCOVA to control for initial differences between the experimental and comparison groups. The contrast between the students’ ITBS pretest and posttest scores was used
to determine the effectiveness of this curriculum in improving their reading comprehension skills.

**Question 3:** A 2 x 2 x 2 factor ANCOVA was used to determine the differential effects of *Jacob's Ladder* by gender and race. The first factor was gender and included the levels of male and female; the second factor of race included the levels of white and non-white. For the purpose of this analysis, students classified as Caucasian were defined as white and students classified as African-American, Asian-American, Hispanic, Native American, or multiracial were classified as non-white.

A 2 x 3 factor ANCOVA was used to determine the differential effects of the curriculum by grade level to include 3rd, 4th, and 5th. The ANCOVAs controlled for any initial statistical significant difference between the two groups when making comparisons between gain scores for each group (Kiess, 2002).

**Question 4:** The teacher and student feedback forms, teacher and student assessment forms, and a random sample of student work collected from teachers at the beginning to the end of implementation were used to analyze the differential effects of the five different genres and four different question ladders included in the *Jacob's Ladder Reading Comprehension Program*. The feedback forms were analyzed to determine the frequency of teachers and students who responded positively to the curriculum versus those who responded negatively. Open-ended questions on these feedback forms were analyzed using the qualitative method of holistic coding (Rossman & Rallis, 2003) to identify emerging patterns of response regarding each genre, ladder type, and/or the curriculum as a whole. The sampling of student products collected from the beginning and end of implementation were analyzed using a paired samples
t-test to determine growth in specific genres and ladder levels over the intervention time frame.

**Question 5:** To determine the potential role played by teacher variability in this study, descriptive statistics were reported for the teacher ratings on the overall COS-R as well as on the six subcategories.

**Question 6:** To determine which critical thinking skills were most enhanced by *Jacob's Ladder*, the mapping of the items on the TCT to Paul's (1992) Reasoning Model and students pre/post gains on the TCT were used to determine significant improvement in specific critical thinking domains. Differences in performance for each of Paul's (1992) domains of critical thinking were then translated into potential differences in the critical thinking domains as determined by the other three critical thinking models used as the foundation of the reading comprehension questions (Appendix F).

Table 5 presents each research question, the corresponding data collected, and statistical analyses employed.
<table>
<thead>
<tr>
<th>Question</th>
<th>Data</th>
<th>Statistical Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How much does the use of Jacob's Ladder enhance the critical thinking</td>
<td>TCT pretest &amp; posttest raw</td>
<td>ANCOVA</td>
</tr>
<tr>
<td>skills of (a) identified gifted learners and (b) identified potentially</td>
<td>scores</td>
<td>Paired samples t-test</td>
</tr>
<tr>
<td>gifted learners?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. How much does the use of Jacob's Ladder improve the reading</td>
<td>ITBS pretest &amp; posttest raw</td>
<td>ANCOVA</td>
</tr>
<tr>
<td>comprehension skills of (a) identified gifted learners and (b) identified</td>
<td>scores</td>
<td>Paired samples t-test</td>
</tr>
<tr>
<td>potentially gifted learners?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. How does the effect of the Jacob's Ladder intervention differ by</td>
<td>TCT pretest &amp; posttest raw</td>
<td>2 x 2 x 2 ANCOVA</td>
</tr>
<tr>
<td>gender, race (white v. non-white), and grade level?</td>
<td>scores</td>
<td>2 x 3 ANCOVA</td>
</tr>
<tr>
<td></td>
<td>ITBS pretest &amp; posttest raw</td>
<td></td>
</tr>
<tr>
<td></td>
<td>scores</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. How does the effect of Jacob's Ladder differ by genre, ladder type,</td>
<td>Teacher &amp; Student Feedback</td>
<td>Descriptive Statistics:</td>
</tr>
<tr>
<td>and ladder level?</td>
<td>Forms</td>
<td>Frequency</td>
</tr>
<tr>
<td></td>
<td>Teacher &amp; Student</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>Assessment Forms</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td></td>
<td>Random Sample of Student</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Products</td>
<td>Holistic coding for emergent</td>
</tr>
<tr>
<td></td>
<td></td>
<td>themes/ patterns</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. How is teacher variability related to student performance on Jacob's</td>
<td>COS-R</td>
<td>Descriptive statistics:</td>
</tr>
<tr>
<td>Ladder tasks?</td>
<td>TCT &amp; ITBS pretest &amp; posttest</td>
<td>Frequency</td>
</tr>
<tr>
<td></td>
<td>scores</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>6. What critical thinking skills are most enhanced by the program?</td>
<td>TCT pretest &amp; posttest</td>
<td>Descriptive statistics:</td>
</tr>
<tr>
<td></td>
<td>scores</td>
<td>Mean domain scores</td>
</tr>
<tr>
<td></td>
<td>Map of TCT items to</td>
<td>Paired samples t-tests</td>
</tr>
<tr>
<td></td>
<td>critical thinking</td>
<td></td>
</tr>
<tr>
<td></td>
<td>domains based on 4 models</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Limitations and Delimitations of the Study

Limitations

The limitations of this study include sample size, differential sampling, history, treatment fidelity, and a lack of generalizability to other districts. The total number of students participating in the study was 154 with 79 experimental students and 75 comparison students. Within the experimental group, 34 students were identified gifted and 45 students were identified as potentially gifted. Within the comparison group, 35 students were identified gifted and 40 were identified as potentially gifted. At each grade level, there were between 20 and 29 students in the experimental classrooms and between 16 and 22 students in the comparison classrooms. Therefore, strong inferences or generalizations cannot be made based on grade level subanalyses because of the small sub-sample sizes. However, any statistical significance found at each grade level indicates a greater power for the intervention than if statistical significance was found with a larger sample.

In addition to sample size, differential sampling posed a threat to validity as well; the groups of potentially gifted low-income students are already participating in one research study and were chosen for this study based on that involvement; these students were randomly assigned to experimental and comparison groups. However, the identified gifted students added to this study were chosen by the school district and were assigned to the experimental or comparison group based on class rosters that were determined before this study was conceptualized. In addition, teachers of the identified gifted students were allowed to choose whether they wanted to participate as an experimental or a comparison group. Therefore, the sampling
procedures used to create these two subgroups were not the same. These differences could potentially alter the results of the data analyses.

When research is conducted in actual classrooms, students’ history can potentially threaten the validity of the study. The researcher cannot control other occurrences in the schools before or while this study is being conducted. Other initiatives or events that happen concurrently could positively or negatively affect the outcome of the curriculum on students’ reading comprehension and critical thinking skills. In this study, concurrent events were a particular problem with the fifth grade experimental students. Because the timing of the intervention fell at the end of the school year, the fifth grade students were involved in many other activities related to their final year of elementary school. In addition, state assessments for all students occurred in the middle of implementation of the Jacob’s Ladder curriculum and immediately prior to the administration of this study’s posttests. The number of tests given to students during this brief amount of time may have affected their motivation and focus when completing the posttests.

Lack of treatment fidelity may have also affected the outcomes of this study. Treatment fidelity is defined as “the extent to which the treatment conditions, as implemented, conform to the researcher’s specifications for treatment” (Gall, Gall, & Borg, 20003, p. 381). The experimental teachers received the implementation guidelines during the training on Jacob’s Ladder prior to beginning implementation. In addition, they all had access to the researcher who, when contacted via email and telephone, provided timely and complete answers to implementation questions on an as-needed basis. There were no restrictions placed on when teachers could contact
the researcher. Classroom observations of all experimental classrooms were also conducted during intervention to ensure treatment fidelity to the greatest extent possible.

Finally, due to the sample size and the scope of the study, the results are generalizable only to other districts with similar characteristics as the primary district in terms of size, average family income, average level of parental education, demographics, and educational practices such as providing services for gifted and high ability students.

**Delimitations**

While *Jacob’s Ladder* may hold promise for all learners, it was developed specifically to meet the needs of high ability and gifted learners involved in a five-year, Javits grant. Therefore, the pilot study has been limited to third, fourth, and fifth grade students who have been identified as gifted or as potentially gifted. It was also developed for use specifically in Title I schools, hence the inclusion of students attending Title I schools. While there are seven districts involved in the larger study from which this research emerged, not all are included in the current intervention. The primary district chosen for the current study was chosen on the basis of geographic proximity and a pre-established relationship with the district, especially the district coordinator of gifted education. Finally, although some researchers might consider the sample size of this study to be a limitation, statistical significance found with a sample size of 154 indicates a more powerful treatment than if statistical significance were found with a larger sample.
Conclusion

This study was designed to test the effectiveness of the Jacob's Ladder Reading Comprehension Program in improving reading comprehension and enhancing critical thinking skills in identified gifted students and identified potentially gifted, low-income students. Through the use of pre- and posttests on a measure of critical thinking and a measure of reading comprehension, the study was intended to determine the extent to which the curriculum improves reading comprehension and critical thinking. The study also hoped to determine whether or not there is a relationship between student performance on the curriculum’s tasks and the variable of the teacher in the classroom. Differential effects by gender, race, and grade level were also investigated as well as whether or not the curriculum is better suited to the improvement of general critical thinking skills or skills in specific critical thinking domains.

The following chapter will address these goals further and will present findings related to each of the six research questions.
CHAPTER IV

Findings

Introduction

This chapter summarizes findings related to each of the six research questions investigating the effects of the Jacob's Ladder Reading Comprehension Program on the reading comprehension and critical thinking skills of identified gifted and potentially gifted students in grades 3, 4, and 5. The study included two treatment and two comparison groups, as outlined in Chapter III; each group was tested before and after the intervention using one measure of reading comprehension and one measure of critical thinking. Additional data were collected from the treatment group to include one classroom observation in each of the 5 experimental classes, student products, teacher assessment of student products, and student and teacher feedback regarding the curriculum. The following sections address findings for each of the research questions and their related data sources.

Findings Related to Question 1

Research Question 1 addressed the extent to which using the Jacob's Ladder Reading Comprehension Program enhanced the critical thinking skills of (a) identified gifted learners and (b) identified potentially gifted learners. This relationship was measured by first analyzing the student gain scores of the treatment and comparison groups for each of the two sample subgroups. Pretest and posttest scores were collected from each of the five classes in the treatment group. Both tests were given in Spring, 2005 with approximately eight to ten weeks between the pretest and the posttest for the identified gifted group and with approximately 16 weeks
between the pretest and posttest for the identified potentially gifted group. It should
be noted that pretest data collection for the fourth grade potentially gifted class
(Classroom 2) was suspect due to unknown testing conditions for the TCT pre-
assessment. The pretest scores on the TCT for the students in this teachers'
classroom were drastically lower than the pretest scores of all other students in the
study indicating a potential problem with test administration. In addition, none of the
students in this classroom completed the pretest which seems to indicate a lack of
sufficient test administration time. Upon conferring with the teacher, it was
discovered that a substitute teacher administered the TCT pretest; therefore, it cannot
be determined whether or not students were given the full 45 minutes to complete the
test. This lack of treatment fidelity with test administration could compromise the
results of the study; therefore, all data analyses involving the TCT scores were
conducted both with and without the student scores from this class. All results of data
analysis for Question 1 are reported both with and without the scores of this class.
The means and standard deviations for students’ scores on the pretest and posttest of
the TCT are reported in Table 6.
Table 6: Means and Standard Deviations of Student Scores on the TCT

<table>
<thead>
<tr>
<th>Sample Group</th>
<th>N</th>
<th>TCT Pretest Mean (SD)</th>
<th>TCT Posttest Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identified Gifted Experimental</td>
<td>34</td>
<td>22.18 (7.79)</td>
<td>25.62 (5.91)</td>
</tr>
<tr>
<td>Identified Gifted Comparison</td>
<td>35</td>
<td>24.25 (4.28)</td>
<td>25.83 (5.53)</td>
</tr>
<tr>
<td>Identified Potentially Gifted Experimental with Classroom 2</td>
<td>45</td>
<td>20.26 (8.23)</td>
<td>26.59 (6.00)</td>
</tr>
<tr>
<td>Identified Potentially Gifted Experimental w/out Classroom 2</td>
<td>32</td>
<td>23.94 (5.89)</td>
<td>26.27 (5.82)</td>
</tr>
<tr>
<td>Identified Potentially Gifted Comparison</td>
<td>75</td>
<td>19.60 (6.05)</td>
<td>24.25 (5.86)</td>
</tr>
</tbody>
</table>

The TCT includes 45 items measuring the Eight Elements of Reasoning as defined by Paul (1992). A raw score of 45 represents a perfect score. An Analysis of Covariance (ANCOVA) was performed to determine treatment effect while controlling for initial differences by covarying the TCT pretest scores. Levene’s Test of Equality of Error Variance was run to ensure equality of variance around the corresponding mean for each group. These tests were non-significant, supporting the use of this statistical procedure.

With an alpha set at .10, significant treatment effects were found for the Identified Potentially Group when the results from Classroom 2 are included. However, when the extreme outlying pretest scores and corresponding posttest scores from Classroom 2 are excluded, no significant treatment effects are evident. Results from the ANCOVA are given in Table 7.
Table 7: ANCOVA Results for Treatment Effects on TCT

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>Between Subjects</th>
<th>( \eta^2 )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identified Gifted</td>
<td>1</td>
<td>(.278)</td>
<td>.111</td>
<td>.002</td>
</tr>
<tr>
<td>( S ) within-group error</td>
<td>65</td>
<td>(25.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identified Potentially Gifted</td>
<td>1</td>
<td>(195.5)</td>
<td>6.928</td>
<td>.056</td>
</tr>
<tr>
<td>( S ) within-group error</td>
<td>116</td>
<td>(28.22)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identified Potentially Gifted w/out Classroom 2 (N=107)</td>
<td>1</td>
<td>(4.15)</td>
<td>.171</td>
<td>.002</td>
</tr>
<tr>
<td>( S ) within-group error</td>
<td>103</td>
<td>(24.30)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Values enclosed in parentheses represent mean square errors.  \( S = \) subjects.  \( p < .10 \)

Because significant treatment effects could not be discerned, a paired samples t-test was run to determine whether or not significant differences from pretest to posttest existed for each group. Significant positive differences between the pretest and posttest scores were found for each group, both with and without the scores from Classroom 2. The effect size for each group was moderate with the identified gifted treatment group showing a somewhat larger effect size than the corresponding comparison group of identified gifted learners. Table 8 presents the results from the paired samples t-test.
Table 8: Results of Paired Samples t-Test for Pre-/Post- Gains on TCT

<table>
<thead>
<tr>
<th></th>
<th>Treatment Group</th>
<th></th>
<th></th>
<th>Comparison Group</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Pretest</td>
<td>Posttest</td>
<td>t</td>
<td>d</td>
</tr>
<tr>
<td>Identified Gifted</td>
<td>34</td>
<td>22.18</td>
<td>25.63</td>
<td>3.16*</td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.79)</td>
<td>(5.91)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identified Potentially</td>
<td>45</td>
<td>20.26</td>
<td>26.59</td>
<td>5.12*</td>
<td>.88</td>
</tr>
<tr>
<td>Gifted</td>
<td></td>
<td>(8.22)</td>
<td>(6.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identified Potentially</td>
<td>32</td>
<td>23.94</td>
<td>26.27</td>
<td>2.66*</td>
<td>.40</td>
</tr>
<tr>
<td>Gifted w/out Classroom 2</td>
<td></td>
<td>(5.89)</td>
<td>(5.82)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .10

Summary of Findings Related to Question 1

The ANCOVA assessed the benefit of using the Jacob’s Ladder Reading Comprehension Program when controlling for pretest differences. No significant results for either the identified gifted group nor the identified potentially gifted group emerged. The paired samples t-test, however, revealed significant positive gains from the TCT pretest to the TCT posttest for all groups, treatment and comparison, identified gifted and identified potentially gifted. The effect sizes for each group were moderate, with the effect size of the identified gifted group being somewhat larger than the effect size for the corresponding group which may suggest some enhancement of critical thinking skills by the curriculum. However, based on the results of the ANCOVA, the gains made by the students on the TCT cannot be attributed to the use of Jacob’s Ladder in their classrooms.

Findings Related to Question 2

Research Question 2 addressed the extent to which using the Jacob’s Ladder Reading Comprehension Program enhanced the reading comprehension skills of (a)
identified gifted learners and (b) identified potentially gifted learners. This relationship was measured by first analyzing the student gain scores on the instrument used to measure reading comprehension for the treatment and comparison groups for each of the two sample subgroups. Pretest and posttest scores were collected from each of the five classes in the treatment group. Both tests were given in Spring, 2005 with approximately eight to ten weeks between the pretest and the posttest for the identified gifted group and with approximately 16 weeks between the pretest and posttest for the identified potentially gifted group.

Students took the Reading portion of the Iowa Test of Basic Skills (ITBS) Survey Battery corresponding to their grade level. All scores were converted to an IQ metric to enable comparison among grade levels. The maximum possible scores are: Grade 3, 147.54; Grade 4, 144.87; and Grade 5, 143.82. The mean scores and standard deviations for each group by grade level are presented in Table 9.

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th></th>
<th>Posttest</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Identified Gifted Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd (N=17)</td>
<td>124.50</td>
<td>9.95</td>
<td>129.32</td>
<td>12.50</td>
</tr>
<tr>
<td>4th (N=17)</td>
<td>125.44</td>
<td>9.96</td>
<td>129.00</td>
<td>10.10</td>
</tr>
<tr>
<td>Identified Potentially Gifted</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd (N=12)</td>
<td>115.88</td>
<td>12.51</td>
<td>122.98</td>
<td>8.74</td>
</tr>
<tr>
<td>4th (N=13)</td>
<td>112.91</td>
<td>7.58</td>
<td>130.19</td>
<td>9.12</td>
</tr>
<tr>
<td>5th (N=20)</td>
<td>120.69</td>
<td>7.38</td>
<td>123.56</td>
<td>8.51</td>
</tr>
<tr>
<td>Comparison Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd (N=19)</td>
<td>126.81</td>
<td>8.27</td>
<td>133.04</td>
<td>11.28</td>
</tr>
<tr>
<td>4th (N=16)</td>
<td>124.30</td>
<td>11.51</td>
<td>129.74</td>
<td>9.84</td>
</tr>
<tr>
<td>5th (N=40)</td>
<td>103.36</td>
<td>13.31</td>
<td>108.02</td>
<td>14.33</td>
</tr>
</tbody>
</table>

Table 9: Mean and SD scores for ITBS by Treatment Group and Grade Level
As the descriptive results in Table 9 show, all students in this study scored well above the average mean of 100 on the ITBS. While only a small percentage of students received a perfect score, the means of several groups were approaching the maximum score. Therefore, the possibility of a ceiling effect should be considered when interpreting the results of further analyses.

An Analysis of Covariance (ANCOVA) was performed to determine treatment effect while controlling for initial differences by covarying the ITBS pretest scores. Levene's Test of Equality of Error Variance was run to ensure equality of variance around the corresponding mean for each group. These tests were non-significant for the identified gifted group, supporting the use of this statistical procedure.

With an alpha set at .10, no significant treatment effects were found for the identified gifted group. Results of the ANCOVA performed are given in Table 10.

<table>
<thead>
<tr>
<th>ITBS-Post (DV)</th>
<th>Df</th>
<th>F</th>
<th>η²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identified Gifted (N=69)</td>
<td>1 (19.73)</td>
<td>.306</td>
<td>.005</td>
<td>.582</td>
</tr>
</tbody>
</table>

Note: Values enclosed in parentheses represent mean square errors. S = subjects. p<.10

For the identified potentially gifted group, the Levene's Test of Equality of Error Variance yielded a significant, F (1, 119) = 18.43, p<.10, indicating that the variances within the experimental and comparison groups were not equal. An ANCOVA was not conducted as the results would not be meaningful; any significant findings would most likely be the result of sampling error.
Because treatment effects were not discernable from the ANCOVA for the identified gifted group and because an ANCOVA could not be performed to determine potential treatment effects for the identified potentially gifted group, a paired samples t-test was run to analyze the differences between pretest and posttest scores on the Reading portion of the ITBS Survey Battery. Significant differences were found between the pretest and posttest scores for all groups. All but one group showed moderate effect sizes; the identified potentially gifted treatment group yielded a large effect size ($d=.88, \eta^2=.402$) indicating approximately 40% of the variance in the ITBS posttest scores can be attributed to belonging to the treatment group. Results of the paired samples t-test are presented in Table 11.

Table 11: Results of Paired Samples t-Test for Pre-/Post- Gains on ITBS

|                      | Treatment Group |                          |                          | Comparison Group |                          |                          |                         |
|----------------------|-----------------|---------------------------|---------------------------|-----------------|---------------------------|---------------------------|                         |
|                      | N               | Pretest M(SD)             | Posttest M(SD)            | t               | d                         | N                        | Pretest M(SD)           | Posttest M(SD)          | T               | D                         |
| Identified Gifted    | 34              | 124.97 (9.81)             | 129.16 (11.19)            | 3.03*           | .40                       | 35                        | 125.66 (9.81)           | 131.53 (10.62)          | 4.11*           | .57                       |
| Identified Potentially Gifted | 45              | 117.16 (9.47)             | 125.32 (9.11)             | 5.27*           | .88                       | 75                        | 113.76 (16.22)          | 118.99 (17.31)          | 5.95*           | .31                       |

* $p<.10$

Summary of Findings Related to Question 2

The ANCOVA used to assess the benefit of using the Jacob's Ladder Reading Comprehension Program yielded no significant results for the identified gifted group. The Levene’s Test of Equality yielded significant results indicating unequal variances in the experimental and comparison groups; this significant finding prohibited the use of an ANCOVA to determine treatment effects for the identified potentially gifted...
A subsequent paired samples t-test revealed significant positive gains from the ITBS pretest to the ITBS posttest for all groups, treatment and comparison, identified gifted and identified potentially gifted. The effect sizes for each group were moderate with the exception of the identified potentially gifted group whose effect size was large. These effect sizes may suggest some enhancement of reading comprehension skills by use of the curriculum, particularly for the identified potentially gifted students. However, based on the results of the ANCOVA, the gains made by the identified gifted students on the ITBS cannot be attributed to the use of *Jacob's Ladder* in their classrooms.

**Findings Related to Question 3**

Question Three addressed whether or not the effects of the *Jacob's Ladder* Reading Comprehension Program differed by gender (male or female), ethnicity (white or non-white), and/or grade level (3rd, 4th, or 5th). Demographic information for each of the students was collected from the TCT and ITBS answer sheets. Using the treatment groups’ TCT and ITBS scores, a 2 x 2 x 2 ANCOVA was run to determine differential effects by gender and by ethnicity. Students who were classified as Caucasian were included in the race category of “white” and students classified as African-American, Asian-American, Hispanic, Native American, or Other were included in the race category of “non-white.” As with Question One, the ANCOVA exploring student gains on the TCT was run both with and without the scores from Classroom 2 and analyses were covaried by the TCT pretest scores. The Levene’s Test of Equality of Error Variances was run with no significant results.
With an alpha set at .10, significant differences were found for the factor of gender on the TCT without the scores from Classroom 2; however, these results should be regarded with caution as the effect size ($\eta^2 = .049$) is small. A subsequent paired samples t-test showed that both males and females made significant gains on the TCT. The effect size for males was moderate while the effect size for females was small. When the scores from Classroom 2 were included, no significant gender differences were found. No significant effects on TCT student gains were found for the factor of ethnicity with either sample. The results of the ANCOVA and the Paired Samples t-test are presented in Tables 12 and 13 respectively.

**Table 12: ANCOVA Results: TCT, Gender x Ethnicity**

<table>
<thead>
<tr>
<th></th>
<th>TCT-Post (DV)</th>
<th>df</th>
<th>F</th>
<th>$\eta^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>With Classroom 2 scores</td>
<td>Gender</td>
<td>1</td>
<td>21.48</td>
<td>.874</td>
<td>.013</td>
</tr>
<tr>
<td></td>
<td>Ethnicity</td>
<td>1</td>
<td>29.97</td>
<td>1.220</td>
<td>.018</td>
</tr>
<tr>
<td></td>
<td>S within-group error</td>
<td>68</td>
<td>24.57</td>
<td></td>
<td>.018</td>
</tr>
<tr>
<td>Without Classroom 2 scores</td>
<td>Gender</td>
<td>1</td>
<td>58.35</td>
<td>3.008</td>
<td>.049</td>
</tr>
<tr>
<td></td>
<td>Ethnicity</td>
<td>1</td>
<td>21.73</td>
<td>1.120</td>
<td>.019</td>
</tr>
<tr>
<td></td>
<td>S within-group error</td>
<td>58</td>
<td>19.40</td>
<td></td>
<td>.019</td>
</tr>
</tbody>
</table>

Note: Values enclosed in parentheses represent mean square errors. $S$ = subjects. *p < .10

**Table 13: Results of Paired Samples t-Test for Gender Effects on TCT Student Gains without Classroom 2**

<table>
<thead>
<tr>
<th></th>
<th>Pretest M (SD)</th>
<th>Posttest M (SD)</th>
<th>Mean Difference (SD)</th>
<th>t</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>23.03 (6.88)</td>
<td>26.58 (5.51)</td>
<td>3.55</td>
<td>3.14*</td>
<td>.57</td>
</tr>
<tr>
<td>Females</td>
<td>23.03 (7.10)</td>
<td>25.26 (6.16)</td>
<td>2.23</td>
<td>2.74*</td>
<td>.34</td>
</tr>
</tbody>
</table>

*p < .10
Similar statistical tests were run to determine the differential effects of gender and ethnicity on student gain scores on the ITBS. With an alpha set at .10, the ANCOVA investigating student gains on the ITBS showed significant differences for the factor of gender, but no significant differences for the factor of ethnicity. A subsequent Paired Samples t-test indicated both males and females made significant gains on the ITBS from pretest to posttest. The effect size for both males and females was moderate. The results of the ANCOVA and Paired Samples t-test are presented in Tables 14 and 15 respectively.

Table 14: ANCOVA Results: ITBS, Gender x Ethnicity

<table>
<thead>
<tr>
<th>ITBS-Post (DV)</th>
<th>df</th>
<th>F</th>
<th>$\eta^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>1</td>
<td>5.001</td>
<td>.069</td>
<td>.029*</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>1</td>
<td>.884</td>
<td>.013</td>
<td>.350</td>
</tr>
<tr>
<td>$S_{within-group error}$</td>
<td>68</td>
<td>(64.93)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Values enclosed in parentheses represent mean square errors. $S =$ subjects. *p<.10.

Table 15: Results of Paired Samples t-Test for Gender Effects on ITBS Student Gains

<table>
<thead>
<tr>
<th>N</th>
<th>Pretest M (SD)</th>
<th>Posttest M (SD)</th>
<th>Mean Difference (SD)</th>
<th>t</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>123.64 (10.15)</td>
<td>128.29 (11.07)</td>
<td>4.65</td>
<td>3.63*</td>
<td>.44</td>
</tr>
<tr>
<td>32</td>
<td>120.30 (10.06)</td>
<td>124.27 (9.16)</td>
<td>3.97</td>
<td>2.68*</td>
<td>.41</td>
</tr>
</tbody>
</table>

*p<.10

Due to the small size of the individual grade level samples, a separate ANCOVA was run to determine the differential effects by grade. For student gain scores on the TCT, the ANCOVA was performed both with and without the scores from Classroom 2. Analyses were covaried by the TCT and the ITBS.
pretest scores. The Levene’s Test of Equality of Error Variances was run with no significant results. No significant differences by grade were found for the student gains on the TCT or the ITBS. The results of these ANCOVAs are presented in Table 16.

<table>
<thead>
<tr>
<th></th>
<th>TCT-Post (DV)</th>
<th>Df</th>
<th>F</th>
<th>( \eta^2 )</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade with Classroom 2 scores</td>
<td></td>
<td>2 (54.69)</td>
<td>2.249</td>
<td>.058</td>
<td>.113</td>
</tr>
<tr>
<td>S within-group error</td>
<td>73 (24.32)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade without Classroom 2 scores</td>
<td></td>
<td>2 (6.53)</td>
<td>.313</td>
<td>.010</td>
<td>.732</td>
</tr>
<tr>
<td>S within-group error</td>
<td>60 (20.84)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITBS</td>
<td></td>
<td>2 (246.68)</td>
<td>3.668</td>
<td>.005</td>
<td>.128</td>
</tr>
<tr>
<td>S within-group error</td>
<td>75 (67.26)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Values enclosed in parentheses represent mean square errors. \( S = \) subjects. \( p<.10 \)

**Summary of Findings from Question 3**

The scores of the experimental group were analyzed to determine if there were any differential effects of the curriculum based on gender, ethnicity, and/or grade level. No significant effects by ethnicity or grade level were found. However, significant differences in student gains on both the TCT and the ITBS were found based on the factor of gender. In both cases, males made significantly larger gains than females, suggesting the curriculum may be of greater benefit to male students than to female students for the enhancement of critical thinking and reading comprehension skills. However, caution should be used when interpreting these results as the effect size was small for the TCT and moderate for the ITBS.
Findings Related to Question 4

The purpose of Question 4 was to determine if the effects of Jacob's Ladder differed based on the genres read (fiction, nonfiction, or poetry), the ladder types completed (A, B, C, or D), and/or the different levels of each ladder (1-3). The data sources used to make this determination included student products, teacher feedback forms, teacher focus groups, and student feedback forms.

Student Products

All five teachers turned in student products from the beginning of the curriculum implementation period; three of the five teachers turned in student products from the end of the implementation period. One of the teachers misplaced her student products from the end of the intervention timeframe while the other teacher forgot to collect the students’ work. Table 17 presents the frequencies for each genre and ladder type included in the student products collected from the treatment group as well as the mean teacher and student scores for each.
### Table 17: Descriptive Statistics for Genre, Ladder Type, and Ladder Level

<table>
<thead>
<tr>
<th>Genre</th>
<th>Freq</th>
<th>Overall Mean (SD)</th>
<th>Level 1 Mean (SD)</th>
<th>Level 2 Mean (SD)</th>
<th>Level 3 Mean (SD)</th>
<th>Teacher Assessment Scores</th>
<th>Overall Mean (SD)</th>
<th>Level 1 Mean (SD)</th>
<th>Level 2 Mean (SD)</th>
<th>Level 3 Mean (SD)</th>
<th>Student Self Assessment Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fiction</td>
<td>165</td>
<td>1.17 (.57)</td>
<td>1.18 (.69)</td>
<td>1.17 (.67)</td>
<td>1.16 (.72)</td>
<td>1.27 (.51)</td>
<td>1.29 (.65)</td>
<td>1.24 (.60)</td>
<td>1.27 (.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poetry</td>
<td>44</td>
<td>1.01 (.62)</td>
<td>1.16 (.81)</td>
<td>.82+ (.67)</td>
<td>1.06 (.80)</td>
<td>1.02 (.63)</td>
<td>1.14 (.64)</td>
<td>1.03 (.65)</td>
<td>.88+ (.69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonfiction</td>
<td>59</td>
<td>1.44 (.59)</td>
<td>1.58 (.72)</td>
<td>1.39 (.79)</td>
<td>1.34 (.73)</td>
<td>1.46 (.48)</td>
<td>1.65 (.55)</td>
<td>1.32 (.58)</td>
<td>1.43 (.57)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>70</td>
<td>1.41 (.52)</td>
<td>1.47 (.60)</td>
<td>1.43 (.67)</td>
<td>1.33 (.70)</td>
<td>1.41 (.47)</td>
<td>1.54 (.53)</td>
<td>1.38 (.58)</td>
<td>1.31 (.64)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>67</td>
<td>1.01 (.61)</td>
<td>1.15 (.74)</td>
<td>1.01 (.69)</td>
<td>.88 (.79)</td>
<td>1.30 (.54)</td>
<td>1.42 (.67)</td>
<td>1.19 (.60)</td>
<td>1.27 (.66)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>94</td>
<td>1.27 (.60)</td>
<td>1.30 (.76)</td>
<td>1.16 (.72)</td>
<td>1.34 (.68)</td>
<td>1.21 (.59)</td>
<td>1.27 (.67)</td>
<td>1.22 (.63)</td>
<td>1.14 (.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>37</td>
<td>.98+ (.57)</td>
<td>.97' (.77)</td>
<td>.91+ (.70)</td>
<td>1.05 (.70)</td>
<td>1.14 (.51)</td>
<td>1.06 (.63)</td>
<td>1.00 (.59)</td>
<td>1.35 (.77)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

+ = Mean score below 1.0

As the descriptive statistics in Table 17 indicate, the fiction was the most popular genre followed by nonfiction and poetry respectively. Ladder C was the most frequently used followed by Ladders A and B with Ladder D being used most infrequently. The mean scores for the majority of the genres and ladders fall between 1.00 and 1.50 which falls within the “Satisfactory” range on the assessment scale for Jacob’s Ladder. Notably, the mean teacher assessment scores for Poetry Level 2, Ladder D overall, and Levels 1 and 2 of Ladder D fell below 1.00, placing these scores in the “Needs Improvement” range. For the mean student assessment scores, the same range applies to Poetry Level 3. Overall, teacher assessment scores and
student assessment scores were similar, with students tending to score themselves slightly higher than their teachers scored them.

Using the mean teacher assessment scores for ladders completed at the beginning of implementation and ladders completed at the end of implementation, independent samples t-tests were run to determine if significant differences existed for any of the three genres or four ladder types. Tests could not be run for the genre of Poetry as only one Poetry ladder completed at the beginning of implementation was included in the student products collected.

With an alpha level set at .10, significant differences between beginning and ending scores in favor of the ending assessment score were found for the Fiction genre, for Ladders A, and Ladder level 3. Significant differences in favor of the beginning score were found for Ladder D; however, this finding should be considered with caution as the ending scores data set consisted of only four student product samples. The results of the t-tests are presented in Table 18.
Table 18: Results of independent samples t-tests of Differential Effects by Genre, Ladder, and Ladder Level

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Beginning M(SD)</th>
<th>Ending M(SD)</th>
<th>t</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fiction</td>
<td>115,50</td>
<td>1.04 (.56)</td>
<td>1.47 (.48)</td>
<td>4.82*</td>
<td>.84</td>
</tr>
<tr>
<td>Nonfiction</td>
<td>36,23</td>
<td>1.52 (.56)</td>
<td>1.30 (.63)</td>
<td>1.36</td>
<td>.36</td>
</tr>
<tr>
<td>Ladder A</td>
<td>43,27</td>
<td>1.31 (.52)</td>
<td>1.56 (.49)</td>
<td>1.99*</td>
<td>.49</td>
</tr>
<tr>
<td>Ladder B</td>
<td>44,23</td>
<td>1.05 (.58)</td>
<td>.94 (.66)</td>
<td>-.708</td>
<td>.18</td>
</tr>
<tr>
<td>Ladder C</td>
<td>32,62</td>
<td>1.16 (.70)</td>
<td>1.33 (.53)</td>
<td>1.31</td>
<td>.27</td>
</tr>
<tr>
<td>Ladder D</td>
<td>33,4</td>
<td>1.03 (.58)</td>
<td>.54 (.25)</td>
<td>-1.66*</td>
<td>1.10</td>
</tr>
<tr>
<td>Level 1</td>
<td>152,116</td>
<td>1.22 (.72)</td>
<td>1.32 (.74)</td>
<td>1.13</td>
<td>.14</td>
</tr>
<tr>
<td>Level 2</td>
<td>152,116</td>
<td>1.12 (.71)</td>
<td>1.22 (.72)</td>
<td>1.10</td>
<td>.14</td>
</tr>
<tr>
<td>Level 3</td>
<td>152,116</td>
<td>1.10 (.74)</td>
<td>1.30 (.72)</td>
<td>2.24*</td>
<td>.28</td>
</tr>
</tbody>
</table>

*p<.10

Teacher Feedback Forms

All five teachers completed and returned feedback forms given to them prior to the end of the intervention period. A copy of the Teacher Feedback form can be found in Appendix C.

Question One asked teachers to describe what aspect of the curriculum worked best in their classrooms. Holistic coding was used to identify emerging themes from teachers' open-ended responses. The themes were overwhelmingly positive with the most often cited benefits being flexibility, ease of implementation, student independence, promotion of higher level thinking, and discussion among students. The emergent codes and their frequencies are listed in Table 19.
Table 19: Emergent Themes from Question 1 on Teacher Feedback Form

<table>
<thead>
<tr>
<th>Emergent Codes</th>
<th>Frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student independence</td>
<td>2</td>
</tr>
<tr>
<td>Ease of implementation</td>
<td>2</td>
</tr>
<tr>
<td>Promotion of higher level thinking</td>
<td>2</td>
</tr>
<tr>
<td>High student interest</td>
<td>1</td>
</tr>
<tr>
<td>Flexibility of curriculum</td>
<td>2</td>
</tr>
<tr>
<td>Quality of literature</td>
<td>1</td>
</tr>
<tr>
<td>Resulting student discussions</td>
<td>2</td>
</tr>
<tr>
<td>Building of language skills</td>
<td>1</td>
</tr>
</tbody>
</table>

Question Two asked teachers to describe the most problematic aspects of implementing *Jacob’s Ladder* in the classroom. No common themes were found among teachers responses to Question Two indicating that difficulties with implementation were classroom-specific. The issues that arose included lack of time, difficulty in facilitating student discussions, insufficient grading system, prohibitive cost of the nonfiction books, Ladder B, and a lack of background information provided for the myths.

Questions Three and Four asked teachers to self-report regarding the number of readings and ladders completed as well as the format(s) used for implementation. The number of readings per genre ranged from two to seven while the number of ladders completed ranged from two to 12. Table 20 presents the teacher reportings for Questions Three and Four by genre and number of ladders completed for each.
Questions Five through Eight asked teachers to provide quantitative data, assessing the clarity of the questions, the clarity of the instructions, the level of student interest, and the perceived educational benefit for students. All teachers rated the curriculum highly on all four questions. The results of these four questions are presented in Table 21.
Table 21: Teacher Responses to Questions 5-8 on Feedback Form

<table>
<thead>
<tr>
<th></th>
<th>Teacher 1</th>
<th>Teacher 2</th>
<th>Teacher 3</th>
<th>Teacher 4</th>
<th>Teacher 5</th>
<th>Overall Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarity of Questions</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2.4</td>
</tr>
<tr>
<td>Clarity of Instructions</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Student Interest</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2.6</td>
</tr>
<tr>
<td>Educational Benefit</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Questions Nine and Ten asked teachers if they would use the program again and if they would recommend it to a fellow teacher. All teachers responded yes to both questions.

Questions Eleven and Twelve asked teachers to rank the genres and the ladders from most to least effective. Fiction was ranked most effective by two teachers, poetry was ranked most effective by two teachers, and nonfiction was ranked most effective by one teacher. Poetry was ranked least effective by two teachers and nonfiction was ranked least effective by three teachers. No teachers ranked fiction as least effective. Teachers were also asked to provide a rationale for their rankings. The rankings and rationales are provided in Table 22.
<table>
<thead>
<tr>
<th>Rankings</th>
<th>Rationale</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Fiction ranked most effective</td>
<td>deep level of textual analysis, high student interest, facilitation of student discussion and notetaking</td>
<td></td>
</tr>
<tr>
<td>Poetry ranked most effective</td>
<td>language building through vocabulary support, creativity</td>
<td></td>
</tr>
<tr>
<td>Nonfiction ranked most effective</td>
<td>(no rationale provided)</td>
<td></td>
</tr>
<tr>
<td>Nonfiction ranked least effective</td>
<td>length of stories, cost of books prohibits the purchase of a class set, tedious</td>
<td></td>
</tr>
<tr>
<td>Poetry ranked least effective</td>
<td>many were difficult to understand, students get enough poetry in the regular curriculum</td>
<td></td>
</tr>
</tbody>
</table>

Three out of five teachers ranked Ladder A as most effective and two teachers ranked Ladder C as most effective. Two teachers ranked Ladder B as least effective, two teachers ranked Ladder A as least effective, and one teacher ranked Ladder C as least effective. The teacher rankings of the ladders and the corresponding rationales are presented in Table 23.
Table 23: Teacher Rankings of Ladders with Corresponding Rationales

<table>
<thead>
<tr>
<th>Rankings</th>
<th>Rationale</th>
</tr>
</thead>
</table>
| Ladder A ranked most effective | • skills relied heavily on the text  
• skills are often tested  
• more often practiced  
• students are more familiar with it |
| Ladder C ranked most effective | • contained the thinking skills least often “tapped” in school  
• inspired the voicing of analytical thought processes |
| Ladder B ranked least effective | • was time consuming  
• students had trouble understanding and creating generalizations  
• the ladder did not provide clarifications of generalizations |
| Ladder A ranked least effective | • “status quo” questioning |
| Ladder C ranked least effective | • Takes more or less the same amount of work as Ladder D  
• Inferences are challenging |

Teacher Focus Groups

Two focus groups were conducted with teachers: one with the two teachers at the Center for the Intellectually Gifted and a separate one with the teachers at the Center for Enrichment. The Teacher Feedback form was used as the structure for each focus group interview. Holistic coding was used to ascertain the emergent themes from these conversations with teachers. Both interviews were overwhelmingly positive and supportive of the curriculum. All five teachers said they would have liked to have more time to use the curriculum in their classrooms, and they felt their students would have benefited from using Jacob's Ladder over the course of a semester or an entire school year. The themes that emerged can be
divided into five categories: Student Response, Instruction/Implementation, Educational Benefit, Future Use of Curriculum, and Improvements to Curriculum. The emergent themes and representative remarks can be found in Appendix H; they are summarized here.

Regarding Student Response to *Jacob's Ladder*, the emergent themes related to student enjoyment of the curriculum, a positive response to student discussion, an appreciation for the shortness of the fiction reading selections, a general dislike for the nonfiction reading selections, the ability of students to relate to the subject matter, and teachers' surprise at students' abilities to grasp abstract concepts.

In the theme category of Instruction/Implementation, teachers specifically discussed students' difficulties with grasping the concepts of generalizations, inferences, and synthesizing; the overly high reading level of some selections in the Level III curriculum; the necessity of directly teaching the terminology associated with each ladder; an appreciation for the flexibility and organization of the curriculum; the benefits of the assessment system in terms of learning about their students; the greater amount of time needed to complete each reading selection compared to their expectations; and the timing of the implementation at the end of the school year as negatively impacting the overall benefit of the curriculum as measured by the research instruments.

Regarding the Educational Benefits of the curriculum, teachers cited the promotion of individual thinking, collaborative thinking, higher level thinking skills, reasoning skills, and student independence as positive aspects of the curriculum.
They also felt that *Jacob's Ladder* sufficiently challenged the students in their classrooms.

In both focus groups, the theme of using *Jacob's Ladder* in the future also emerged. All teachers had ideas for how they would use the curriculum to its greatest advantage in future years; these ideas included using *Jacob's Ladder* for the entire school year, for the second marking period, and to recommending the curriculum to other teachers.

Finally, in terms of how *Jacob's Ladder* could be improved, the teachers involved in the focus groups gave the following ideas: add more readings/ladders to each genre; vary the reading level within each curriculum level to a greater extent to incorporate a wider range of reading abilities; change the way the nonfiction readings are approached; and change the answer sheet.

**Student Feedback**

Overall, 75 students completed and returned the Student Feedback form. A copy of this form can be found in Appendix C. The Student Feedback form does not specifically ask students about individual genres and ladder types. However, it does ask students to respond to the overall curriculum in terms of enjoyment, challenge, and educational benefits. Students were asked to respond with a “yes” or a “no” to Questions One through Four and to provide a rationale for their answers. For Questions One, Three, and Four, the majority of students answered “yes”. For Question Two, regarding the enjoyment and benefit of student discussions, the students’ answers were fairly evenly split between “yes” and “no”. While many of the students said they enjoyed the program because it was fun, many others
recognized the benefits of discussing literature with other students, challenging questions, going back to the text to answer questions, persistence, and learning to read better. Students who did not enjoy the program said, for the most part, that it was either too boring or too hard; or they simply stated that they do not like to read. Students who responded negatively to the discussions cited reasons such as being shy, not agreeing with other students' answers, and being afraid of getting the answer wrong in front of their peers. The total number of “yes” and “no” answers for each question as well as illustrative rationales given by students can be found in Appendix I.

Question Five asked students to give advice to the authors on how they could improve the curriculum while question Six asked for any additional comments. Several students said the program was great as is; others thought there should be more stories and questions. The additional comments included extending the ladders to four levels and thanking the authors for the learning experience. Table 24 presents student responses to questions Five and Six.
Table 24: Response to Student Feedback Questions 5 and 6

<table>
<thead>
<tr>
<th>Question 5: Advice to authors for improvement</th>
<th>“I don’t think you have to improve the program because it’s just fine.”</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Put in real stories</td>
<td>• Add poems, nonfiction books, and yearbooks</td>
</tr>
<tr>
<td>• Add poems, nonfiction books, and yearbooks</td>
<td>• Make more questions</td>
</tr>
<tr>
<td>• Make more questions</td>
<td>• Make it more fun</td>
</tr>
<tr>
<td>• Make it more fun</td>
<td>• Put in a new section with rhymes</td>
</tr>
<tr>
<td>• Put in a new section with rhymes</td>
<td>• Make some of the stories funny</td>
</tr>
<tr>
<td>• Make some of the stories funny</td>
<td>• Add more stories</td>
</tr>
<tr>
<td>• Add more stories</td>
<td>• Write about more things we can relate to like “My Sister is a Sissy”</td>
</tr>
<tr>
<td>• Write about more things we can relate to</td>
<td>• “My Shadow”</td>
</tr>
<tr>
<td>• Make the stories longer</td>
<td>• Add more myths and folktales</td>
</tr>
<tr>
<td>• Use less questions</td>
<td>• Add harder questions</td>
</tr>
<tr>
<td>• Add more science</td>
<td>• Put stories about kids in it</td>
</tr>
<tr>
<td>• Add more science</td>
<td>• Make the stories longer</td>
</tr>
<tr>
<td>• Add fantasy stories with dragons</td>
<td>• Use less questions</td>
</tr>
<tr>
<td>• Make it less poetic</td>
<td>• Add more science</td>
</tr>
<tr>
<td>• Add more about sports</td>
<td>• Add more challenge</td>
</tr>
<tr>
<td>• Add more historical stories</td>
<td>• “Take it easy on the questions—some were very, VERY hard.”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question 6: Additional Comments</th>
<th>It was interesting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I love doing your poems</td>
</tr>
<tr>
<td></td>
<td>Thank you for the learning experience</td>
</tr>
<tr>
<td></td>
<td>Add some 4 ladders like B1, B2, B3, B4 for 5th grade</td>
</tr>
</tbody>
</table>

Summary of Findings Related to Question 4

The analysis of student products showed room for growth in all three genres and all four ladders, with mean scores well below the maximum score of 2.00.

Students showed significant progress on questions related to fiction readings, on Ladder A, and on Level Three of all ladders included in the analysis. Students appeared to regress on Ladder D, but these results may not be reliable as the sample of student products completed at the end of implementation consisted of only four for
this ladder. These results suggest moderate differential effects for genre, ladder type, and ladder level in favor of fiction, Ladder A, and Level 3.

The analysis of Teacher Feedback forms and the comments made during teacher focus groups indicate an overwhelmingly positive response to most aspects of the Jacob's Ladder curriculum. The majority of teachers were not as pleased with the nonfiction readings as they were with the fiction and poetry readings mostly because of the time required to complete a nonfiction reading with the corresponding ladders and because of a lack of sufficient materials. All ladders were considered effective, with no one ladder taking a clear lead as most effective over the other ladders.

The analysis of the Student Feedback forms revealed that most students responded positively to the curriculum. For three out of the four "yes" or "no" questions the majority of students responded that they enjoyed the readings in the program, that they learned from the program, and that they would like to use the program again next year. The responses to the question regarding student discussion were more equally split between "yes" and "no". There were comparable numbers of students who enjoyed talking about their work versus students who would rather not share their work with others.

**Findings Related to Question 5**

Question 5 focused on the effects of teacher variability on student performance on the critical thinking and reading comprehension skills targeted by Jacob's Ladder. Each treatment group teacher was observed once during the intervention period and the Classroom Observation Scale-Revised (COS-R) was completed for each. The maximum Mean score is 3.0 for the overall COS-R and for
each of its subcategories of Curriculum Planning and Delivery, Accommodating for Individual Differences, Problem Solving, Critical Thinking Strategies, Creative Thinking Strategies, and Research Strategies. A Treatment Fidelity form was also completed to ensure all teachers were implementing the curriculum as intended; the maximum score possible on the Treatment Fidelity form was a 9 with one point being given for an observed intended behavior and no points given for not observed behaviors. Table 25 presents the scores on the COS-R and Treatment Fidelity form for each teacher.

<table>
<thead>
<tr>
<th>Classroom Observation Scale-Revised (COS-R)</th>
<th>Teacher 1</th>
<th>Teacher 2</th>
<th>Teacher 3</th>
<th>Teacher 4</th>
<th>Teacher 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>2.73</td>
<td>1.18</td>
<td>2.60</td>
<td>2.44</td>
<td>2.78</td>
</tr>
<tr>
<td>CPD</td>
<td>3.00</td>
<td>1.00</td>
<td>2.67</td>
<td>2.67</td>
<td>2.50</td>
</tr>
<tr>
<td>AID</td>
<td>2.67</td>
<td>1.50</td>
<td>3.00</td>
<td>3.00</td>
<td>2.67</td>
</tr>
<tr>
<td>PS</td>
<td>Not observed</td>
<td>Not observed</td>
<td>Not observed</td>
<td>Not observed</td>
<td>Not observed</td>
</tr>
<tr>
<td>CRIT</td>
<td>2.5</td>
<td>1.00</td>
<td>2.50</td>
<td>2.00</td>
<td>3.00</td>
</tr>
<tr>
<td>CREA</td>
<td>2.67</td>
<td>1.00</td>
<td>2.50</td>
<td>2.00</td>
<td>3.00</td>
</tr>
<tr>
<td>RS</td>
<td>Not observed</td>
<td>Not observed</td>
<td>Not observed</td>
<td>Not observed</td>
<td>Not observed</td>
</tr>
<tr>
<td>Treatment Fidelity</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

As the descriptive statistics presented in Table 25 demonstrate, four of the five teacher scores were comparable, ranging from mean scores of 2 (somewhat effective) to 3 (effective) on the COS-R and each of its subcategories. One teacher score was significantly lower than the other four teacher scores with mean scores between 1.00 (ineffective) and 1.50. None of the teachers exhibited behaviors in the Problem Solving or the Research Strategies categories. All five teachers scored a seven out of
nine on the Treatment Fidelity form indicating that all teachers were implementing the curriculum as intended. The most common behaviors not observed in the classrooms included: differentiating instruction based on student strengths/weaknesses (4), students discussing literature as a whole group (3), students discussing readings in dyads (1), and students completing record sheets (1).

However, it is not expected that all nine behaviors on the Treatment Fidelity form would be observed during one classroom observation; therefore, a score of seven out of nine on the Treatment Fidelity form strongly supports evidence of appropriate implementation of the curriculum by all teachers. However, the teacher in Classroom 2, despite all efforts, did not challenge her students to the extent desired. This teacher allowed students to copy each other’s answers, to turn in their work without self-assessing their answers, and to complete entire ladders without input from the teacher. This teacher’s rating on the COS-R reflects her lack of effectiveness as compared to the other four teachers.

Summary of Findings Related to Questions 5

The classroom observations conducted during the intervention period yielded results on the COS-R which show some variability among teachers, most notably Teacher 2 whose scores varied widely from the other four teachers. Based on their scores on the Treatment Fidelity form, all five teachers implemented the curriculum in their classrooms as intended.
Findings Related to Question 6

Question 6 focused on determining what critical thinking skills are most enhanced by using the Jacob's Ladder Reading Comprehension Program. The 45 items on the TCT have been mapped to Paul's (1992) Elements of Reasoning; this alignment is presented in the TCT manual (Bracken et al., 2005). An equal number of TCT items focus on each of the Elements. Table 26 presents the eight Elements of Reasoning and the corresponding TCT items.

Table 26: Paul's Elements of Reasoning and the Corresponding TCT items

<table>
<thead>
<tr>
<th>Elements</th>
<th>Corresponding TCT Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Issue</td>
<td>2, 4, 5, 20, 22, and 34</td>
</tr>
<tr>
<td>Purpose</td>
<td>8, 13, 23, 28, and 40</td>
</tr>
<tr>
<td>Concept</td>
<td>7, 9, 14, 26, and 31</td>
</tr>
<tr>
<td>Point of View</td>
<td>3, 12, 16, 30, 33, and 39</td>
</tr>
<tr>
<td>Assumption</td>
<td>15, 19, 27, 36, 41, and 44</td>
</tr>
<tr>
<td>Evidence</td>
<td>11, 18, 29, 35, 38, and 43</td>
</tr>
<tr>
<td>Inference</td>
<td>1, 6, 10, 17, 24, and 42</td>
</tr>
<tr>
<td>Implication</td>
<td>21, 25, 32, 37, and 45</td>
</tr>
</tbody>
</table>

To determine which elements were most enhanced by using the curriculum, a Multivariate Analysis of Variance (MANOVA) was performed on the TCT pretest to posttest differences for each of the eight Elements of Reasoning based on treatment condition. MANOVA's were run both with and without the scores from Classroom 2. Levene's Test of Equality of Variance was run for each MANOVA with no significance found.

With an alpha level set at .10, significant pre-/post- differences were found for the Elements of Purpose and Point of View both with and without the scores from
Classroom 2. No significant differences were found for the other six Elements.

Table 27 presents the results of the MANOVAs run to determine the effects of using *Jacob's Ladder* on specific critical thinking skills based on treatment group.

**Table 27: MANOVA Results: TCT pre/post Gains by Treatment Group on Elements of Reasoning**

<table>
<thead>
<tr>
<th>Issue</th>
<th>With Classroom 2 Scores</th>
<th>Without Classroom 2 Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Df 1 (.28) F .165 p .685 $\eta^2$ .001</td>
<td>Df 1 (.09) F .054 p .816 $\eta^2$ .000</td>
</tr>
<tr>
<td>Purpose</td>
<td>1 (9.40) F 6.952 p .009 $\eta^2$ .047</td>
<td>1 (10.77) F 8.125 p .005 $\eta^2$ .058</td>
</tr>
<tr>
<td>Concept</td>
<td>1 (.08) F .068 p .795 $\eta^2$ .000</td>
<td>1 (.19) F .151 p .698 $\eta^2$ .001</td>
</tr>
<tr>
<td>POV</td>
<td>1 (8.19) F 4.730 p .031 $\eta^2$ .032</td>
<td>1 (9.83) F 5.645 p .019 $\eta^2$ .041</td>
</tr>
<tr>
<td>Assumption</td>
<td>1 (.41) F .197 p .658 $\eta^2$ .001</td>
<td>1 (4.7) F .223 p .638 $\eta^2$ .002</td>
</tr>
<tr>
<td>Evidence</td>
<td>1 (2.55) F 2.346 p .128 $\eta^2$ .016</td>
<td>1 (1.86) F 1.713 p .193 $\eta^2$ .013</td>
</tr>
<tr>
<td>Inference</td>
<td>1 (3.49) F 2.285 p .133 $\eta^2$ .016</td>
<td>1 (1.65) F 1.030 p .312 $\eta^2$ .008</td>
</tr>
<tr>
<td>Implication</td>
<td>1 (1.80) F .876 p .351 $\eta^2$ .006</td>
<td>1 (2.66) F 1.290 p .258 $\eta^2$ .010</td>
</tr>
</tbody>
</table>

Note: Values enclosed in parentheses represent mean square errors. $p<.10$

The results in Table 27 indicate which of Paul's (1992) Elements of Reasoning appear to be most enhanced by using *Jacob's Ladder*. However, these Elements of Reasoning are not the only critical thinking skills targeted by the curriculum. Appendix G presents a mapping of Paul's (1992) Elements to the critical thinking skills included in Bloom's Taxonomy, Taba's Concept Development Model, and the Creative Problem Solving Model. Using this mapping and the results from Table 28, inferences can be made about other critical thinking skills enhanced by the program. The critical thinking skills mapped to Paul's (1992) Element of Purpose include Bloom's levels of Remembering, Analyzing, Evaluating, and Creating; Taba's levels of generating Details and formulating Generalizations; and the Creative Problem Solving skills of Mess Finding and Problem Finding. The critical thinking

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
skills mapped to Paul’s (1992) Element of Point of View include Bloom’s levels of Understanding, Applying, and Analyzing; Taba’s levels of generating details and the classification or categorization of those details; and the Creative Problem Solving skills of Mess Finding, Data Finding, Problem Finding, and Idea Finding. Table 28 presents the skills from each of the thinking models included in the curriculum as well as their relationship to the curriculum and to the two Paul’s (1992) Elements of Reasoning that were significantly affected by the use of Jacob’s Ladder.

Summary of Findings Related to Question 6

The MANOVA intended to show the effects of using Jacob’s Ladder on specific critical thinking skills showed significant results for the Elements of Purpose and Point of View suggesting that students in the experimental group made significantly greater gains on TCT items relating to these two critical thinking elements. Inferences were then made to include skills from the three other thinking models included in the design of Jacob’s Ladder. Although the critical thinking domains included in these models are not directly tested by the TCT, a mapping of Paul’s (1992) Elements of Reasoning to the specific skills targeted by the other thinking models was developed. This mapping was used to show the relationship among the Elements of Purpose and Point of View to many of the skills integrated in Bloom’s Taxonomy, Taba’s Concept Development Model, and the Creative Problem Solving model. However, it should be noted that the validity of the mapping among the thinking strategies has not been assessed by external raters.
Table 28: Mapping of Thinking Models, Paul's Elements of Purpose & Point of View, Jacob's Ladder Ladder & Level

<table>
<thead>
<tr>
<th></th>
<th>Paul's Element of Purpose</th>
<th>Paul's Element of POV</th>
<th>Jacob's Ladder Ladder &amp; Level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bloom's Taxonomy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remembering</td>
<td>X</td>
<td></td>
<td>B1</td>
</tr>
<tr>
<td>Understanding</td>
<td></td>
<td>X</td>
<td>C1</td>
</tr>
<tr>
<td>Applying</td>
<td></td>
<td>X</td>
<td>C1</td>
</tr>
<tr>
<td>Analyzing</td>
<td></td>
<td>X</td>
<td>B3</td>
</tr>
<tr>
<td>Evaluating</td>
<td></td>
<td>X</td>
<td>B3</td>
</tr>
<tr>
<td>Creating</td>
<td></td>
<td>X</td>
<td>B3</td>
</tr>
<tr>
<td>Details</td>
<td></td>
<td>X</td>
<td>B1</td>
</tr>
<tr>
<td><strong>Taba's Concept</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Development</td>
<td>Classification/</td>
<td>X</td>
<td>D2</td>
</tr>
<tr>
<td></td>
<td>Categorization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generalizations</td>
<td></td>
<td>X</td>
<td>B3</td>
</tr>
<tr>
<td>Mess Finding</td>
<td></td>
<td>X</td>
<td>B1, C1</td>
</tr>
<tr>
<td>Data Finding</td>
<td></td>
<td></td>
<td>C1</td>
</tr>
<tr>
<td>Problem Finding</td>
<td></td>
<td>X</td>
<td>D2</td>
</tr>
<tr>
<td>Idea Finding</td>
<td></td>
<td>X</td>
<td>D2</td>
</tr>
<tr>
<td>Solution Finding</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Acceptance Finding</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

**Summary of Findings**

The research findings from this study of the Jacob’s Ladder Reading Comprehension Program fall into two major categories: findings based on the quasi-
experimental aspect of the study as it relates to the effects of the intervention and those findings based on feedback from students and teachers.

Key Findings Related to Effects of the Intervention

1. The ANCOVA intended to assess the benefit of using the Jacob's Ladder Reading Comprehension Program yielded no significant results for either the identified gifted group or the identified potentially gifted group. The Paired Samples t-test revealed significant positive gains from the TCT pretest to the TCT posttest for all groups, treatment and comparison, identified gifted and identified potentially gifted. The effect sizes for each group were moderate with the effect size of the treatment identified gifted group being somewhat larger than the effect size for the corresponding group which may suggest some enhancement of critical thinking skills by the curriculum. However, based on the results of the ANCOVA, the gains made by the students on the TCT cannot be attributed to the use of Jacob's Ladder in their classrooms.

2. The ANCOVA intended to assess the benefit of using the Jacob's Ladder Reading Comprehension Program yielded no significant results for the identified gifted group. A significant interaction was found between membership in the identified potentially gifted group and the covariate of the ITBS pretest scores; this significant finding prohibited the use of an ANCOVA to determine treatment effects for the identified potentially gifted group. A subsequent Paired Samples t-test revealed significant positive gains from the ITBS pretest to the ITBS posttest for all groups, treatment and comparison, identified gifted and identified potentially gifted. The effect sizes
for each group were moderate with the exception of the treatment identified potentially gifted group whose effect size was large. These effect sizes may suggest some enhancement of reading comprehension skills by use of the curriculum, particularly for the identified potentially gifted students. However, based on the results of the ANCOVA for the identified gifted group, the gains made by the students on the ITBS cannot be attributed to the use of Jacob's Ladder in their classrooms.

3. The scores of the experimental group were analyzed to determine if there were any differential effects of the curriculum based on gender, ethnicity, and/or grade level. No significant effects by ethnicity or grade level were found. However, significant differences in student gains on both the TCT and the ITBS were found based on the factor of gender. In both cases, males made significantly larger gains than females suggesting the curriculum may be of greater benefit to male students than to female students for the enhancement of critical thinking and reading comprehension skills. However, caution should be used when interpreting these results as the effect size was small for the TCT and moderate for the ITBS.

4. The MANOVA intended to show the effects of using Jacob's Ladder on specific critical thinking skills showed significant results for the Elements of Purpose and Point of View, suggesting that students in the experimental group made significantly greater gains on TCT items relating to these two critical thinking elements. Inferences were then made to include skills from the three other thinking models included in the design of Jacob's Ladder. Although the
critical thinking domains included in these models are not directly tested by the TCT, a mapping of Paul’s (1992) Elements of Reasoning to the specific skills targeted by the other thinking models was developed. This mapping was used to show the relationship among the Elements of Purpose and Point of View to many of the skills integrated in Bloom’s Taxonomy, Taba’s Concept Development Model, and the Creative Problem Solving model. However, it should be noted that the validity of the mapping among the thinking strategies has not been established.

5. The classroom observations conducted during the intervention period yielded results on the COS-R which show some variability among teachers, most notably Teacher 2 whose scores varied widely from the other four teachers. Based on their scores on the Treatment Fidelity form, all five teachers implemented the curriculum in their classrooms as it was intended. However, the limited variability among teacher scores, the small sample size, the relatively short length of intervention, and the limited scope of one classroom observation should be taken into consideration when interpreting these results.

6. The analysis of student products showed room for growth in all three genres and all four ladders, with mean scores well below the maximum score of 2.00. Students showed significant progress on questions related to fiction readings, on Ladder A, and on Level Three of all ladders included in the analysis. Students appeared to regress on Ladder D, but these results may not be reliable as the sample of student products completed at the end of implementation consisted of only four for this ladder. These results suggest
moderate differential effects for genre, ladder type, and ladder level in favor of fiction, Ladder A, and Level 3.

Summary of Findings Related to Teacher and Student Feedback

1. The analysis of Teacher Feedback forms and the comments made during teacher focus groups indicate an overwhelmingly positive response to most aspects of the Jacob's Ladder curriculum. The majority of teachers were not as pleased with the nonfiction readings as they were with the fiction and poetry readings, mostly because of the time required to complete a nonfiction reading with the corresponding ladders and because of a lack of sufficient materials. All ladders were considered effective with no one ladder taking a clear lead as most effective over the other ladders.

2. The analysis of the Student Feedback forms revealed that most students responded positively to the curriculum. For three out of the four “yes” or “no” questions, the majority of students responded that they enjoyed the readings in the program, that they learned from the program, and that they would like to use the program again next year. The responses to the question regarding student discussion were more equally split. There were comparable numbers of students who enjoyed talking about their work versus students who would rather not share their work with others.

The following chapter discusses these findings with respect to the literature on critical thinking skills and reading comprehension instruction. Some conclusions will be drawn based on the results and some suggestions for implications as well as for further research and practice will be made.
CHAPTER V
Discussion, Conclusions, and Implications

Introduction

The purpose of this study was to pilot the *Jacob's Ladder Reading Comprehension Program* with identified gifted and potentially gifted learners in grades 3, 4, and 5. The following areas of focus were of particular interest: the effects of this curriculum on students' critical thinking and reading comprehension skills; how the use of this curriculum was affected by teacher variability; and what specific organizational aspects of the curriculum were most and least effective. All of the instruments used in the study were intended to address one or more of the focal areas as well as to analyze the relationship between students' demographic information and their response to the curriculum.

The results pertaining to these focal areas, as presented in Chapter IV, will be discussed in the following section. The goal of this chapter is to elucidate the conclusions that can be drawn from this research, to explore the implications of this study for future research and practice, and to discuss unexpected findings from this applied study.

The Effects of Using *Jacob's Ladder*

*Jacob's Ladder* was developed to enhance both the critical thinking and reading comprehension skills of high ability learners. To explore the effectiveness of the curriculum in these three areas, this study used several instruments to explore the effects of *Jacob's Ladder* on students' critical thinking skills and their reading comprehension skills as well as the effects of teacher variability on student
performance related to the curricular tasks. The current study also explored whether or not the effects of using *Jacob's Ladder* would be different for students identified as gifted versus students identified as potentially gifted.

**The Effects on Critical Thinking Skills**

Critical thinking is a difficult concept to study because it is difficult to characterize. Many different definitions of critical thinking exist and each study that explores this concept does so in a different manner (Stahl, 1991). However, researchers and practitioners agree that critical thinking embodies invaluable skills for students to learn (Dixon, 2002; Gallagher, 1998; Little, 2002; Paul, 1991; Thompson, 2002), and efforts continue to design curriculum to develop such skills. Previous research on curriculum designed to teach critical thinking has shown that critical thinking skills can be systematically taught in the classroom (Halpern, 1994; Lipman, 2003). The curriculum that has been shown effective shares many characteristics with the *Jacob's Ladder* curriculum such as an emphasis on inferences, analysis, synthesis, multiple points of view, independent learning, modeling, logical sequencing of tasks, reflection, and adequate support for teachers (Halpern, 1994; Lipman, 2003). In addition, research on using Paul's (1992) Model of Reasoning, which served as the conceptual framework for *Jacob's Ladder*, to enhance critical thinking skills have yielded positive results (VanTassel-Baska and Bracken, 2005; VanTassel-Baska, Little, Rogers, Feng, & Drummond, 2002). However, none of the results from the current study justified similar conclusions for students identified as gifted or for students identified as potentially gifted. The analysis of pretest to posttest gains on the Test of Critical Thinking both between and among comparison
groups was the primary data source for determining the effects of *Jacob's Ladder* on students’ critical thinking skills. The results of the pretest-posttest analysis, as detailed in Chapter IV, showed that students in both the treatment and the comparison groups appeared to improve their critical thinking regardless of identification status. On a descriptive level, the experimental group showed larger gains than the comparison group; however, the difference between student gains in each group was not statistically significant. Therefore, the conclusion cannot be made that improvement of the critical thinking skills of students in the treatment group was the result of participation in the intervention. Several possible explanations for the nonsignificant results exist.

When considering the quantitative analyses using the liberal \( p \) value of .10, a reasonable conclusion would be that the curriculum may not be an especially effective means of enhancing students’ critical thinking skills. However, further research on the use of the curriculum would be advisable before drawing this conclusion. The strongest support for further research regarding the curriculum’s effects on students’ critical thinking skills is the testimony of teachers and students in support of the curriculum. All of the teachers and the majority of students using the curriculum felt it was a positive addition to their classroom. Teachers especially reported an improvement in students’ higher level thinking skills after being exposed to the readings and questions included in *Jacob's Ladder*. All the teachers and many of the students stated they would like to use *Jacob's Ladder* in future years. In addition, although the treatment group’s scores on the TCT did not show significantly more improvement than the comparison group’s scores, they did show some
improvement. The data, both qualitative and quantitative, support the premise that using this curriculum does not negatively impact student learning. Therefore, additional research using Jacob's Ladder would not harm students and may yield positive results if the lessons learned from the current study are addressed.

Furthermore, although significant results were not found for students' overall gains on the TCT, significant effects were found for specific subsets of the TCT related to the eight different domains of Paul's (1992) Reasoning Model. The TCT Manual (Bracken et al., 2005) maps each of the items on the TCT to one of the Elements of Reasoning defined by Paul (1992). Subanalyses of these items reveal that students in the experimental group made significantly greater gains in the Reasoning Elements of Purpose and Point of View. These particular critical thinking skills are targeted by three of the four ladder skill sets included in the Jacob's Ladder Reading Comprehension curriculum. Therefore, although the results of the statistical analyses performed in this study do not definitively show enhancement of critical thinking skills overall, through the use of this curriculum, informal data and subanalyses on the TCT show positive support for its continued use and for its potential benefits over a longer period of time.

One possible factor to explain the lack of student gains attributable to the use of Jacob's Ladder may be the length of the intervention in this study. The studies showing the effective use of Paul's (1992) Reasoning Model involved interventions that ranged in length from five months to three years (VanTassel-Baska & Bracken, 2005; VanTassel-Baska, Avery, Hughes, & Little, 2000; VanTassel-Baska, Hughes, Avery, & Little, 1996; VanTasssel-Baska, Little, Rogers, Feng, & Drummond, 2002;
VanTassel-Baska, Zuo, Avery, & Little, 2002); the current study employed an implementation period of eight weeks. While attempts were made to increase the length of the intervention, competing demands for teachers’ time and the increasing number of expectations placed on schools prevented such efforts. In their feedback regarding *Jacob’s Ladder*, teachers cited the lack of time as a hindrance to the benefits of the curriculum in their classrooms. Most teachers felt that the curriculum needed to be used for at least three-quarters of the year in order to fulfill its potential.

Related to the short length of intervention and another variable to consider is the amount of time necessary for a person to develop critical thinking skills and the prerequisite existence of critical thinking dispositions (Beyer, 1985; Cacioppo & Petty, 1982, 1983; DeNitto & Strickland, 1987; Facione, 1991; Halpern, 1989, 1997; MacDonald, 1970; Paul & Nosich, 1991; Taube, 1995). The development of critical thinking dispositions is considered imperative for critical thinking because, according to the literature, a person without motivation, perseverance, and a desire to actively engage in seeking appropriate solutions to problems cannot be a critical thinker (Beyer, 1985; Cacioppo & Petty, 1982, 1983; DeNitto & Strickland, 1987; Facione, 1991; Halpern, 1989, 1997; MacDonald, 1970; Paul & Nosich, 1991; Taube, 1995). Specific dispositions considered necessary for successful critical thinking include an awareness of the need to evaluate information; a willingness to test assumptions; a desire to consider all viewpoints (Beyer, 1985); independence of thought; fairmindedness, intellectual humility; intellectual courage; confidence in reason (Paul & Nosich, 1991); a tolerance for ambiguity (MacDonald, 1970); and an ability to discriminate between strong and weak arguments (Cacioppo & Petty, 1982, 1983).
The acquisition of these skills could take several years or more. If students participating in this study did not already possess the necessary critical thinking dispositions before using Jacob's Ladder, it is unlikely they were able to develop these dispositions and subsequently improve their critical thinking skills within the eight week intervention period.

Another variable that may have influenced the outcome of this analysis was the timing of beginning the intervention at the end of the school year. The fifth grade students in particular were involved in numerous activities to celebrate and commemorate their final weeks of elementary school that could have been academically distracting. Previous research has shown the need for at least moderate intensity free from distraction for the development of critical thinking strategies, especially metacognition (Barton & Sawyer, 2003; Chiu, 1998). In addition, the posttesting for the current study occurred three to seven days after the administration of state assessment tests. It is possible that students were test wary and therefore lacked the motivation and focus they exercised when taking the pretest. This lack of motivation and focus could contribute to a depression of posttest scores. Comparison students in third and fourth grade experienced similar posttesting conditions, but the comparison students in fifth grade did not. The scores for the fifth grade comparison group were gathered earlier in the school year; therefore, their posttest scores may not have been negatively affected by competing demands for their attention. Given the small sample size, this variation in testing conditions could have contributed to the lack of a significant difference between student gains in the experimental versus the comparison group.
The Effects on Reading Comprehension Skills

Previous research on reading comprehension instruction has supported the use of multiple types of questions with varying degrees of complexity, of graphic organizers, and of generic visual aids over an extended period of time (Anderson & Biddle, 1975; Armbruster et al., 1987; Duke & Pearson, 2002; Levin & Pressley, 1981). Teacher modeling of effective comprehension strategies has also been found effective as have the strategies of comprehension monitoring, cooperative learning, discussion, locating details, comparing/contrasting, determining cause and effect, drawing conclusions, making thematic connections, taking multiple perspectives, and metacognition (Barton & Sawyer, 2003; Chiu, 1998; Guthrie & Anderson, 1999; Kolic-Vehovec & Bajsanski, 2001; Warian, 2003). All of these characteristics of reading comprehension instruction are included in the Jacob's Ladder Reading Comprehension curriculum. Therefore, based on evidence from prior research, the curriculum used in this current study should improve students' reading comprehension skills. However, as described in Chapter IV, the results of the pretest-posttest analysis on the measure used to determine reading comprehension growth, the ITBS, are inconclusive.

As with the TCT, both the experimental and the comparison group made significant gains on the ITBS Reading Survey Battery. However, for the identified gifted group, a statistically significant difference between experimental and comparison group gains was not found. For the identified potentially gifted group, comparisons could not be made between the student gains in the experimental group and student gains in the control group because of a lack of equality in sample
variance between the experimental and comparison groups. Based on the standard
deviations for each groups' mean scores, it is likely that this interaction was due to
the large variance within the comparison group. Because the comparison group was
created by combining students from two different districts, this variability is not
surprising despite every effort being made to choose a secondary district that was
similar to the primary district involved in this study.

When looking at the quantitative data and considering the liberal $p$ value of
.10, a logical explanation that must be considered for the lack of significant treatment
effects is that Jacob's Ladder is not an effective curriculum for enhancing the reading
comprehension skills of identified gifted or identified potentially gifted learners.
However, the same positive, qualitative data previously used to support further
research on the curriculum's ability to enhance critical thinking skills applies to the
enhancement of reading comprehension skills; both teachers and students felt the
reading selections, the need for re-reading the text, and the requirement of text-based
support for answers were all beneficial educational practices.

A possible contributing variable to the lack of significant findings with the
reading comprehension analysis of the identified gifted students' scores was the
potential for ceiling effect on the ITBS. All the students involved in this study scored
well above average on the pretest, with some students receiving a perfect score.
Given that the experimental group was identified as gifted with mean IQ scores at the
85 percentile or higher, this should not be too surprising. Due to the lack of room for
growth on this measure, enhancements to students' reading comprehension skills
have been undetectable with this instrument. An off-level achievement test would
have provided more useful data for assessing the reading comprehension skills of these students. However, in an effort to minimize the number of additional assessments given to students, the decision was made to use the same instruments as those being used in Project Athena, especially considering that the research on *Jacob's Ladder* is an extension of the larger Athena study.

As with the results of the TCT analysis, another possible explanation for the lack of significant differences between student gains in the identified gifted experimental group and in the corresponding comparison group could be the length of intervention and the timing of the intervention. The same distractions that may have affected student performance on the posttest of the TCT could also apply to student performance on the ITBS posttest.

An interesting finding that emerged from the subanalyses by gender for both the TCT and the ITBS was that statistically significant differences were found for male students versus female students. For both instruments, male students appeared to make greater improvements in critical thinking and reading comprehension skills than female students. From these findings, several conclusions can be drawn.

Based on the findings of these subanalyses, the curriculum may be of greater benefit for male students than for female students. With respect to critical thinking skills, gender differences have not been noted in the literature or in previous research. No gender differences were found in the reviewed research on developing effective curriculum to teach critical thinking (Halpern, 1994; Lipman, 2003) or in the reviewed research on curriculum using Paul's (1992) Reasoning Model (VanTassel-
Baska and Bracken, 2005; VanTassel-Baska, Avery, Hughes, & Little, 2000; VanTassel-Baska, Hughes, Avery, & Little, 1996; VanTasssel-Baska, Little, Rogers, Feng, & Drummond, 2002; VanTassel-Baska, Little, Rogers, Feng, & Drummond, 2002). Therefore, if the gender differences discovered in the current study are true differences, they are most likely due to some aspect of the curriculum design such as the visual representation of the ladders moving students from lower level to higher level questions which may be supported by research findings that males are more spatial learners (Bonanno & Kommer, 2005; Colom, Contreras, Arend, Leal, & Santacreu, 2004; Rilea, Roskos-Ewoldsen, & Boles, 2004). In addition, scaffolding is a commonly used strategy in instruction for students with learning disabilities of which a greater percentage are males (Bender, 2002; Hagen, 1991; Liederman, Kantrowitz, & Flannery, 2005). However, due to the size of the current sample, the lack of significant treatment effects, and the limited generalizability of the study findings, any conclusions made about differential benefits of the curriculum favoring male learners would be speculative.

Student products could provide another data source for exploring the potential differences in the effects of Jacob's Ladder based on gender. However, demographic information for the random selection of student products collected for this sample was not requested nor provided. Therefore, statistical analyses cannot be run to determine if male students made greater gains than female students on performance on curriculum-specific tasks. The inclusion of such data analyses in future research using Jacob's Ladder would be beneficial in clarifying the potential existence of gender differences.
A probable explanation for the gender differences noted is variance within the small sample of experimental students. The small effect sizes for the differences in the pretest-posttest scores on both the TCT and the ITBS indicate that only a small amount of the variance in scores can be attributed to the factor of gender.

*The Effects of Teacher Variability*

Research on teacher effectiveness has consistently shown that teacher behavior significantly affects student learning, and that ineffective teachers have a negative impact on student performance whereas effective teachers positively impact student learning (Sanders & Horn, 1998; Sanders & Rivers, 1996; Mendro, 1998; Stronge, 2002; Tucker & Stronge, 2005). This assertion appears to be true for the current study. Four of the five teachers involved in the study were rated comparably on the COS-R. However, one teacher, the teacher in Classroom 2, was rated considerably lower on the COS-R than her colleagues. Classroom 2 is the same classroom in which the TCT pretest data collection errors occurred and in which the observer noted a lack of high expectations and challenging demands for students as they progressed through the ladders.

Analyses to determine the statistical effects of teacher variability could not be run as a sample size of five teachers is too small to produce reliable statistical results in a correlation analysis. It should also be noted that the ratings on the COS-R involve a “snapshot” of the classroom and the teachers’ instructional behaviors. One classroom observation was conducted in each of the experimental classrooms, capturing approximately 45 minutes of instructional time. These 45 minutes represent only a small portion of the learning activities taking place in the classrooms.
at any given time. Each of the teacher ratings on the COS-R must be considered within the limited context of one classroom observation. Similarly, one person completed each of the classroom observations; an error in judgment in rating the teachers could potentially skew the results of the correlation analysis.

The Effects of Organizational Aspects of the Curriculum

There are several unique design features of the Jacob's Ladder Reading Comprehension curriculum that were explored in this study, including the reading selection genres, the variety of skill ladders, the design of the ladders with three different levels of question difficulty, and the inclusion of student discussion. The decision to include these particular design features was based on research on instructional practices for gifted learners and for reading comprehension (Knapp et al., 1995; Taylor et al., 2003; VanTassel-Baska, French, & Stambaugh, 2004). The results of the current study offered some preliminary indications that these design features are effective for student learning; other results, though, were inconclusive.

Reading Selection Genres

With respect to the reading selection features, the overall results were somewhat inconclusive due to a lack of sample student work in the genre of poetry. Feedback from teachers and students indicated an enjoyment of the poetry selections with several of the teachers, and many students responded that the poetry was their favorite part of the curriculum. However, student progress in poetry could not be determined quantitatively because of the random process through which student work samples were collected. While teachers were asked to try to provide samples from each of the three genres, all of the teachers focused more on providing student
samples from the beginning and end of implementation in the areas of fiction and nonfiction.

Student scores on tasks related to nonfiction reading selections were lower at the end of the implementation period compared to their scores at the beginning of implementation, at a descriptive level. The decrease in their scores, however, was not statistically significant. Teacher and student feedback was least positive for the nonfiction genre than for any other genre included in the curriculum. While several students responded positively to the subject matter covered in the nonfiction books, all of the teachers felt these books were too long, the activities took too much time, and the lack of sufficient books for the entire class impeded student progress. Because of the time factor, teachers asked students to complete fewer nonfiction ladders than fiction and poetry ladders. All of these issues could have contributed to teachers' and students' spending less time completing nonfiction reading selections and ladders which in turn could affect student performance scores. Similarly, student focus on the tasks could have wavered as the time required to complete the task demands increased thereby leading to a decrease in student scores from the beginning of implementation when all experiences were new to the end of implementation when students were more familiar with the curriculum.

Analysis of student scores within the fiction genre yielded statistically significant results for student scores at the end of implementation as compared to their scores at the beginning of implementation, indicating positive student progress. Student scores at the end of implementation were still well below the maximum score possible showing sufficient room for continued student growth. Of all the student
samples collected, the majority were related to fiction reading selections; however, based on teacher reports, fiction was not the primary focus of instruction with Jacob's Ladder. Teachers reported completing as many or more selections in poetry as in fiction. The positive results found regarding student scores on student work samples, therefore, warrant more systematic data collection and analysis of student work in the other genres, particularly poetry.

Ladder Types and Levels

An equivalent number of student samples were collected across the four different ladder types with the exception of the fourth ladder, Ladder D, which focuses on paraphrasing, summarizing, and creative synthesis. While a similar number of student answers to the fourth ladder were collected at the beginning of implementation, only four samples were collected at the end of implementation, making statistical analysis of student progress on this ladder invalid. The only other ladder to yield statistically significant results was the first ladder, Ladder A, which focuses on sequencing, cause and effect, and consequences and implications. While the results for Ladder A are encouraging, the overall results for the organizational feature of the ladders are inconclusive. While feedback from teachers indicate that all four ladders are beneficial to student learning, teachers appeared more impressed with the variety of reading selections and the increasing difficulty of questions as students moved up the ladders.

Findings related to the different ladder levels were also somewhat inconclusive, but did yield positive results for student improvement on ladder level three, the highest, most complex, and most abstract level of each skill ladder. From a
descriptive standpoint, students showed improvement on all three levels of the ladders. These findings suggest that the scaffolding created through the visual representation of climbing a ladder as the questions become more demanding is an effective strategy for student learning; they also support the findings from previous research on the use of scaffolding in reading comprehension instruction (Fielding & Pearson, 1994; Villaume & Brabham, 2002).

*Student Discussion*

In this study, teachers reiterated prior research claims that providing students with opportunities to discuss their work and receive feedback improved the students’ reading comprehension responses to literature (Chin et al., 2001; Pressley et al., 2001; Taylor et al., 2002). Many of the teachers’ positive comments about Jacob’s Ladder focused on student discussion as the aspect that contributed most to student learning and to teacher awareness of students’ understanding. However, the student feedback regarding student discussion was more mixed. A relatively equal number of students stated that they enjoyed the discussions as those who indicated they did not enjoy the discussions. The students who appreciated the discussion component cited reasons such as liking to hear other students’ answers, getting a chance to talk to other students about literature, and getting to know other students better. Those students who did not enjoy the discussion component cited reasons such as it was a waste of time, no one could ever agree on a “right” answer, they did not like to share their answers with others, and they preferred to work alone. Perhaps the differences in student response to the discussion aspect of Jacob’s Ladder can be attributed to different personalities and different learning styles. Or, perhaps, they can be
attributed to a difference in teacher implementation of the curriculum and the approaches used in the classroom to promote discussion.

Implications for Research and Practice

Implications for Research

As a pilot study for the Jacob's Ladder Reading Comprehension Program, this study presented few findings of statistical significance regarding the effects of the curriculum on identified gifted and identified potentially gifted students' critical thinking and reading comprehension skills. As an exploratory study, however, this investigation did provide information about the key questions as well as several directions for further research warranted by the findings.

First, Jacob's Ladder needs to be given longer than eight weeks in the classroom before definitively answering questions regarding its effectiveness in enhancing students' critical thinking and reading comprehension skills. Existing research and literature suggests the need for a semester or longer of moderate intensity instruction and practice to improve these skills (Barton & Sawyer, 2003; Chiu, 1998; Van-Tassell-Baska & Bracken, 2005; VanTassel-Baska, Avery, Hughes, & Little, 2000; VanTassel-Baska, Hughes, Avery, & Little, 1996; VanTasssel-Baska, Little, Rogers, Feng, & Drummond, 2002; VanTassel-Baska, Zuo, Avery, & Little, 2002). All of the teachers involved in this study agree with these assertions; they stated that Jacob's Ladder is a worthwhile curriculum that should be used for longer periods of time ranging from three-quarters to a full year. In addition, descriptive statistics seem to indicate a larger improvement in experimental students'
performance that might have reached statistical significance over a greater length of time.

The timing of the intervention and assessment should also be carefully planned. *Jacob's Ladder* may be more beneficial to students if it is presented at the beginning of the school year rather than at the end. Many activities are competing for teacher and student attention at the end of the school year, making it a less than ideal time to present new material. All of the teachers in this study said they would rather begin using the curriculum during the first semester. The timing of testing is also important. With the proliferation of standardized testing to assess student learning, students are spending more time than ever taking tests. Future intervention studies using *Jacob's Ladder* should consider scheduling pre- and post-assessments during times that are not already being used for other standardized tests.

Consideration should also be given to using off-level tests for the measurement of reading comprehension skills. Given students' pretest scores, ceiling effect may be a threat when grade level tests are used. Since the ITBS measures similar skills of increasing difficulty across grade levels, giving third graders the fourth grade test, fourth graders the fifth grade test, and fifth graders the sixth grade test may provide a more accurate representation of student improvement in reading comprehension.

The discontinuation of a standardized measure of critical thinking should also be considered. It is difficult to show short term gains on standardized measures of critical thinking (Cacioppo & Petty, 1982, 1983; DeNitto & Strickland, 1987; Ennis, 1993; Facione, 1991); the development of critical thinking skills is a lifelong process.
In addition, the need for preexisting critical thinking dispositions makes it difficult to determine if a lack of growth in critical thinking is due to a lack of curriculum efficacy or insufficient development of these dispositions in students. Alternate means of determining student growth, such as performance-based assessment of the specific skills targeted by the Jacob's Ladder curriculum, should be considered.

The question of the effects of teacher variability also presents some interesting directions for further research. The literature and research on teacher effectiveness suggests that more effective teachers would lead to better student performance (Darling-Hammond, 2004a, 2004b; Sanders & Rivers, 1996; Mendro, 1998; Schalock, 1998; Stronge, 2002; Tucker & Stronge, 2005; Wright, Horn, & Sanders, 1997). Given the research support for the power of the teacher (Darling-Hammond, 2004a, 2004b; Sanders & Rivers, 1996; Mendro, 1998; Schalock, 1998; Stronge, 2002; Tucker & Stronge, 2005; Wright, Horn, & Sanders, 1997), additional exploration of the question of teacher variability should be conducted with a larger sample size that would facilitate a statistical correlation analysis and multiple classroom observations.

Research on professional development would also suggest a more structured approach to teacher training (Borko, Elliott, & Uchiyama, 2002; Borko, Mayfield, Marion, Flexer, & Cumbo, 1997; Westberg, Archambault, Dobyns, & Salvin, 1993; Westberg & Daoust, 2003). Borko et al. (1997, 2002) recommend eliminating one-time professional development sessions that are not beneficial for teachers. Instead, professional development and training should be ongoing; should engage teachers at their current interest and skill levels; should assess and address teacher beliefs about
the targeted instructional strategies; and should be placed within the context of the classroom (Borko, Elliott, & Uchiyama, 2002; Borko, Mayfield, Marion, Flexer, & Cumbo, 1997). Using this literature and research-base, the training for future Jacob's Ladder interventions should include a hands-on, interactive professional development session, a initial attempt to implement the curriculum in the classroom, and a follow-up session with all teachers to revisit the implementation guidelines while simultaneously addressing teachers' questions or concerns.

A more systematic approach to obtaining student products is also recommended based on the findings from the current study. While the student samples analyzed were informative, more reliable conclusions could have been drawn if products had been obtained from each genre and ladder in equal numbers. Additionally, looking at student work towards the middle of implementation might also yield important information regarding student progress throughout their experience with Jacob's Ladder. Furthermore, asking students to include demographic information with their work samples may help determine if there are truly gender effects associated with the curriculum through an analysis of student performance data by gender. The development of performance-based assessments focused on the skills most directly targeted by Jacob's Ladder would also facilitate the collection of data specific to students' performance on the curriculum tasks. The use of performance-based assessment to facilitate the improvement of critical thinking and literary analysis skills is supported by previous research on language arts curriculum for high ability learners (VanTassel-Baska, Avery, Hughes, & Little,
In addition, the results of this study did not suggest any differential effects of *Jacob's Ladder* based on student status as identified gifted or identified potentially gifted. Teacher feedback suggests that all students can benefit from the targeted scaffolding in reading comprehension and critical thinking. Therefore, additional studies exploring the use of *Jacob's Ladder* with a wider range of student abilities may be warranted.

Finally, the issue of the setting for future studies utilizing *Jacob's Ladder* should be considered. As Chapters I and III described, *Jacob's Ladder* was developed with the intention of providing teachers with a curriculum they could easily implement in their already overly scheduled school days. Therefore, there is no alternative but to continue exploration of the curriculum's effectiveness in real classroom settings with real classroom conditions. However, the difficulties with data collection from one classroom, the lack of sufficient intervention time, and the difficulty with finding a demographically comparable comparison group may have all contributed to the dearth of significant treatment effects associated with the use of *Jacob's Ladder*. These experiences are informative for future researchers exploring the use of this curriculum in real classrooms, especially with respect to communicating early with districts and teachers, scheduling of assessment, and interacting frequently with teachers in their classrooms.
Implications for Practice

In terms of practice, teacher and student feedback indicate that Jacob's Ladder was a welcome addition to their classrooms. All of the teacher responses and the majority of student responses indicated a desire to use the curriculum again in future classes. Suggestions were made, however, for the improvement of the curriculum. Most notably, teachers felt that dividing the nonfiction reading selections into sections and providing question ladders for each section would make this genre more manageable for instruction. In addition, several teachers requested a more detailed assessment system to facilitate the translation of scores on Jacob's Ladder tasks to traditional letter grades required by school districts. Teachers would also like to begin using the curriculum earlier in the school year and more completely integrate the reading selections into the core language arts curriculum.

Based on student responses, the discussion component of Jacob's Ladder was the only aspect that did not receive overwhelmingly positive comments. The discussion, however, is critical to enhancing students' critical thinking and reading comprehension skills with this curriculum. Therefore, a more systematic approach to teacher modeling and organizing of discussion may need to be developed for the professional development of teachers on using this curriculum. Perhaps if students had a greater understanding of the purpose of discussion and a specific discussion format to follow, they would respond more positively to the experience.

Summary

Critical thinking and reading comprehension skills are critical to student success in the language arts classroom. Enhancement of these skills is a worthwhile
endeavor for curriculum developers to undertake. This study was intended to explore
the effectiveness of Jacob's Ladder, a curriculum designed to enhance such skills for
identified gifted and identified potentially gifted students in grades 3, 4, and 5.

While the results of this study did not definitively support the effectiveness of
Jacob's Ladder, the findings indicate positive movement in the direction of enhanced
learning. Teacher and student response to the curriculum was overwhelmingly
positive, and the findings from specific content analyses of ladder elements suggest
potential benefits of using this curriculum. As an exploratory study, this research
provided numerous directions for further, more in-depth investigations of the use of
Jacob's Ladder as an intervention program to enhance reading comprehension and
critical thinking.
References


Bloom, B.S. (1956). *Taxonomy of educational objectives: The classification of"
Bonanno, P., & Kommers, P.A.M. (2005). Gender differences and styles in the use of

Kentucky's educational reform effort. *Teaching and Teacher Education, 18*,
969-987.

developing ideas and practices: Successes, stumbling blocks, and implications

Hieronymous, A.; Frisbie, D.; & Dunbar, S. (Eds.), *Mental Measurements
Yearbook* [Electronic version]. Retrieved January 14, 2005, from
http://spweb.silverplatter.com.proxy.wm.edu/c2430

Brushaber, T. (2003). Teaching comprehension through a comprehension strategy
framework. (ERIC Document Reproduction Service No. ED477160)


Cacioppo, J.T., Petty, R.E., & Morris, K.J. (1983). Effects of the need for cognition
on message evaluation, recall, and persuasion. *Journal of Personality and Social
Psychology, 45*, 805-818.

Metuchen, NJ: Scarecrow Press.


modifications for students who are gifted. Paper presented at the meeting of American Council on Rural Special Education, Albuquerque, NM. (ERIC Document Reproduction Service No. ED429750)


Gubbins, E.J. (2002). *Implementing a professional development model using gifted"
education strategies with all students (RM02172). Storrs, CT: The National Research Center on the Gifted and Talented, University of Connecticut.


Heath, W. J. (1997). What are the most effective characteristics of teachers of the gifted? (ERIC Document Reproduction Services No. ED411665)


Hoy, A.W., & Spero, R.B. (2005). Changes in teacher efficacy during the early years


National Reading Panel (2000a). Teaching children to read: An evidence-based


Nordlund, M. (2003). Differentiated instruction: Meeting the educational needs of all students in your classroom. Lanham, MD: ScarecrowEducation.


Parke, B.N. (1995). Developing curricular interventions for the gifted. In Genshaft,
J.L., Bireley, M., & Hollinger, C.L. (Eds.), Serving gifted and talented students: a resource for school personnel (pp.123-134). Austin, TX: Pro-Ed.


Stahl, N. N., & Stahl, R. J. (1991). We can agree after all! achieving consensus for a
critical thinking component of a gifted program useing the delphi technique.

_Roeper Review, 14(2), 79-88._


*University of Iowa: Iowa Testing Program* (n.d.) Retrieved on January 25, 2005, from


observational study of instructional and curricular practices used with gifted and talented students in regular classrooms (RM93104). Storrs, CT: The National Research Center on the Gifted and Talented, University of Connecticut.


APPENDIX A:

Classroom Observation Scale-Revised
The William and Mary Classroom Observation Scales, Revised
Teacher Observation
Joyce VanTassel-Baska, Ed.D. Linda Avery, Ph.D. Jeanne Struck, Ph.D.

**Directions:** Please employ the following scale as you rate each of the checklist items. Rate each item according to how well the teacher characteristic or behavior was demonstrated during the observed instructional activity. Each item is judged on an individual, self-contained basis, regardless of its relationship to an overall set of behaviors relevant to the cluster heading.

<table>
<thead>
<tr>
<th>3=Effective</th>
<th>2=Somewhat Effective</th>
<th>1=Ineffective</th>
<th>N/O = Not Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>The teacher evidenced careful planning and classroom flexibility in implementation of the behavior, eliciting many appropriate student responses. The teacher was clear, and sustained focus on the purposes of learning.</td>
<td>The teacher evidenced some planning and/or classroom flexibility in implementation of the behavior, eliciting some appropriate student responses. The teacher was sometimes clear and focused on the purposes of learning.</td>
<td>The teacher evidenced little or no planning and/or classroom flexibility in implementation of the behavior, eliciting minimal appropriate student responses. The teacher was unclear and unfocused regarding the purpose of learning.</td>
<td>The listed behavior was not demonstrated during the time of the observation. (NOTE: There must be an obvious attempt made for the certain behavior to be rated &quot;ineffective&quot; instead of &quot;not observed&quot;).</td>
</tr>
</tbody>
</table>

**General Teaching Behaviors**

<table>
<thead>
<tr>
<th>Curriculum Planning and Delivery</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>N/O</th>
</tr>
</thead>
<tbody>
<tr>
<td>The teacher...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. set high expectations for student performance.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. incorporated activities for students to apply new knowledge.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. engaged students in planning, monitoring or assessing their learning.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. encouraged students to express their thoughts.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. had students reflect on what they had learned.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Differentiated Teaching Behaviors**

<table>
<thead>
<tr>
<th>Accommodations for Individual Differences</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>N/O</th>
</tr>
</thead>
<tbody>
<tr>
<td>The teacher...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. provided opportunities for independent or group learning to promote depth in understanding content.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. accommodated individual or subgroup differences (e.g., through individual conferencing, student or teacher choice in material selection and task assignments.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. encouraged multiple interpretations of events and situations.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. allowed students to discover key ideas individually through structured activities and/or questions.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Problem Solving**

<table>
<thead>
<tr>
<th>3</th>
<th>2</th>
<th>1</th>
<th>N/O</th>
</tr>
</thead>
<tbody>
<tr>
<td>The teacher...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. employed brainstorming techniques.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
11. engaged students in problem identification and definition
12. engaged students in solution-finding activities and comprehensive solution articulation.

Comments:

**Critical Thinking Strategies**

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>N/O</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.</td>
<td>encouraged students to judge or evaluate situations, problems, or issues</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td>engaged students in comparing and contrasting ideas (e.g., analyze generated ideas)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>provided opportunities for students to generalize from concrete data or information to the abstract</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>encouraged student synthesis or summary of information within or across disciplines</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comments:

**Creative Thinking Strategies**

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>N/O</th>
</tr>
</thead>
<tbody>
<tr>
<td>17.</td>
<td>solicited many diverse thoughts about issues or ideas</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.</td>
<td>engaged students in the exploration of diverse points of view to reframe ideas</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19.</td>
<td>encouraged students to demonstrate open-mindedness and tolerance of imaginative, sometimes playful solutions to problems</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.</td>
<td>provided opportunities for students to develop and elaborate on their ideas</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comments:

**Research Strategies**

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>N/O</th>
</tr>
</thead>
<tbody>
<tr>
<td>(It is atypical for these to be observed in one session. Some teachers, however, may use items #21-25 within a single period to illustrate the full research process to students. Please note those observations in the comments section.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.</td>
<td>required students to gather evidence from multiple sources through research-based techniques (e.g., print, non-print, internet, self-investigation via surveys, interviews, etc.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22.</td>
<td>provided opportunities for students to analyze data and represent it in appropriate charts, graphs, or tables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23.</td>
<td>asked questions to assist students in making inferences from data and drawing conclusions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.</td>
<td>encouraged students to determine implications and consequences of findings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25.</td>
<td>provided time for students to communicate research study findings to relevant audiences in a formal report and/or presentation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comments:

Additional Comments:
APPENDIX B:

Treatment Fidelity Form
**Jacob’s Ladder Treatment Fidelity Form**

Teacher: ___________________________  Date: ________________  Observation #: ______

<table>
<thead>
<tr>
<th>Implementation</th>
<th>Observed</th>
<th>Not Observed</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students complete initial answers individually</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students are grouped in dyads for discussion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students are completing self-evaluations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students are completing record sheets</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher is differentiating reading selections based on student strengths and weaknesses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher is providing student feedback</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students are completing reading selections from each genre</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students are discussing literature as a whole group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students and teacher are conferring on readings</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX C:

Teacher and Student Feedback Forms
Teacher Feedback Form

Please take a few minutes to share your comments about the Jacob's Ladder Program. Your feedback will be used to enhance the program.

1. What aspect of Jacob's Ladder worked best in the classroom?

2. What was most problematic about implementing Jacob's Ladder?

3. How many readings and ladders did your students complete for each genre?

<table>
<thead>
<tr>
<th>Genre</th>
<th>Readings</th>
<th>Ladders</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonfiction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fables/Myths</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poems</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. How did you implement the program? (check all that apply)
   - Centers
   - Student dyads
   - Whole class work
   - Small group work with whole class discussion
   - Independent work for selected students
   - Selected reading group
   - Other: Please explain:

5. How would you assess the clarity of the questions in the ladders? 3 2 1

6. How would you rate the clarity of instructions for use? 3 2 1

7. How would you rate student interest while using the program? 3 2 1

8. How would you assess the educational value of the program for your students? 3 2 1
9. Would you use the program again?
   If no, why not? ____________________________ □ yes □ no

10. Would you recommend the program to a fellow teacher?
    If no, why not? ____________________________ □ yes □ no

11. Please rank the reading selection genres from most effective to least effective.

<table>
<thead>
<tr>
<th>Most effective</th>
<th>1st ranked genre:</th>
</tr>
</thead>
<tbody>
<tr>
<td>2nd ranked genre:</td>
<td></td>
</tr>
<tr>
<td>3rd ranked genre:</td>
<td>Least effective</td>
</tr>
</tbody>
</table>

Please provide a rationale below for your ranking of genres.

12. Please rank the questions ladders (Ladder A, B, C, or D) from most effective to least effective.

<table>
<thead>
<tr>
<th>Most effective</th>
<th>1st ranked ladder:</th>
</tr>
</thead>
<tbody>
<tr>
<td>2nd ranked ladder:</td>
<td></td>
</tr>
<tr>
<td>3rd ranked ladder:</td>
<td></td>
</tr>
<tr>
<td>4th ranked ladder:</td>
<td>Least effective</td>
</tr>
</tbody>
</table>

Please provide a rationale below for your ranking of ladders.

Other comments:

Thank you for your time and all your hard work on this project.
Student Feedback Form

Please take a few minutes to let us know your reactions to the Jacob's Ladder Program.
Thank you. The Authors

1. Did you enjoy the readings in the program? Why or why not? □ yes □ no

2. Did you enjoy talking about the questions? Why or why not? □ yes □ no

3. Do you feel you learned important things from the program? If yes, list what you feel were the three most important things you learned.
   a. ______________________
   b. ______________________
   c. ______________________ □ yes □ no

4. Would you like to do more activities like this next year? If no, why not? □ yes □ no

5. What advice do you have for the authors in how to improve the program?

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Do you have anything else you would like to tell us about your experience with the program? (You may continue your answer on the back.)
APPENDIX D:

Standards Alignment Chart Example
# PILOT STUDY OF JACOB'S LADDER

## Level III: Standards Alignment

<table>
<thead>
<tr>
<th>Language Arts - Essays/Short Stories</th>
<th>The Gettysburg Address</th>
<th>Brazilian Paradise</th>
<th>Christa McAuliffe</th>
<th>Common Sense</th>
<th>The Competition</th>
<th>Legacy</th>
<th>Moving Pictures Evoke Concern</th>
<th>FDR Inaugural Address</th>
<th>Washington's Letter</th>
<th>Why Own a House</th>
</tr>
</thead>
<tbody>
<tr>
<td>The student will use analysis of text, including the interaction of the text with reader's feelings and attitudes to create response.</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>The student will interpret and analyze the meaning of literary works from diverse cultures and authors by applying different critical lenses and analytic techniques.</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>The student will integrate various cues and strategies to comprehend what he or she reads.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>The student will use a knowledge of the purposes, structures, and elements of writing to analyze and interpret various types of text.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
## Level III: Standards Alignment

<table>
<thead>
<tr>
<th>Language Arts - Essays/Short Stories</th>
<th>The Gettysburg Address</th>
<th>Brazilian Paradise</th>
<th>Christa McAuliffe</th>
<th>Common Sense</th>
<th>The Competition</th>
<th>Legacy</th>
<th>Moving Pictures Evoke Concern</th>
<th>FDR Inaugural Address</th>
<th>Washington's Letter</th>
<th>Why Own a House</th>
</tr>
</thead>
<tbody>
<tr>
<td>The student will use a knowledge of the purposes, structures, and elements of writing to analyze and interpret various types of text.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
PILOT STUDY OF *JACOB'S LADDER*

APPENDIX E:

Ladder A-D Templates
PILOT STUDY OF JACOB'S LADDER

Generalizations

Classifications

Details

Title of Reading Selection:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
PILOT STUDY OF JACOB'S LADDER

Theme/Concept

Evidence/Inference

Characterization

Title of Selection:

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

Mapping of Paul’s Reasoning Model to Bloom’s Taxonomy, Taba’s Concept Development Model, Creative and Problem Solving
PILOT STUDY OF *JACOB'S LADDER*

<table>
<thead>
<tr>
<th>Model</th>
<th>Component</th>
<th>Ladder A</th>
<th>Ladder B</th>
<th>Ladder C</th>
<th>Ladder D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1 2 3</td>
<td>1 2 3</td>
<td>1 2 3</td>
<td>1 2 3</td>
</tr>
<tr>
<td>Bloom's Taxonomy - Updated</td>
<td>Remembering</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td></td>
<td>Understanding</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td></td>
<td>Applying</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td></td>
<td>Analyzing</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td></td>
<td>Evaluating</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td></td>
<td>Creating</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Taba's Concept Develop.</td>
<td>Details</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td></td>
<td>Classification/ Categorization</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td></td>
<td>Generalizations</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Creative Problem Solving</td>
<td>Mess Finding</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td></td>
<td>Data Finding</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td></td>
<td>Problem Finding</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td></td>
<td>Idea Finding</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td></td>
<td>Solution Finding</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td></td>
<td>Acceptance Finding</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
</tbody>
</table>
APPENDIX G:

Implementation Guide for Teachers
Implementation Procedures for *Jacob's Ladder* Reading Comprehension Program

<table>
<thead>
<tr>
<th>Week of Implementation</th>
<th>Implementation Guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week One</td>
<td>• Administer the TCT following the instructions provided in the manual</td>
</tr>
<tr>
<td></td>
<td>• Administer the Reading subtest of the ITBS following the instructions provided in the manual</td>
</tr>
<tr>
<td></td>
<td>• Begin using <em>Jacob's Ladder</em> using the first reading selection in the first genre section of your notebook; work through the corresponding skill ladders as a class. You should model the following process:</td>
</tr>
<tr>
<td></td>
<td>o Read the selection independently</td>
</tr>
<tr>
<td></td>
<td>o Independently answer the questions in the corresponding skill ladders beginning at the bottom rung and “climbing up”</td>
</tr>
<tr>
<td></td>
<td>o Share answers to each question with you partner; begin at the bottom rung and work your way up</td>
</tr>
<tr>
<td></td>
<td>o Discuss your answers with your partner; try to reach consensus on a best answer for each question</td>
</tr>
<tr>
<td></td>
<td>o Fill out the student self-assessment and the individual reading selection assessment</td>
</tr>
<tr>
<td></td>
<td>o If desired and time permits, pairs of students may share their answers with the rest of the class</td>
</tr>
<tr>
<td></td>
<td>o Staple individual answers, any changes made during discussion, and all assessment sheets together; hand in to teacher</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
### PILOT STUDY OF JACOB'S LADDER

<table>
<thead>
<tr>
<th>Week One (con’t)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Ensure all students understand the process of Jacob's Ladder by providing an opportunity for a Q&amp;A session. If necessary, repeat the modeling process for all of some students.</td>
<td></td>
</tr>
<tr>
<td>• Assign students the first reading selection in the second genre section. Have students complete the process on their own. The independent reading and completion of the skill ladder questions can be completed in class or as a homework assignment.</td>
<td></td>
</tr>
<tr>
<td>• Students should complete two Jacob's Ladder reading selections (beyond the one modeled during class), the corresponding ladder sets and assessment sheets by the end of the first week. During the discussion process, you should circulate among groups providing feedback as needed to keep the discussion on track and an appropriately challenging level.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Week Two</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Using the student assessment sheets as a guide, determine the strengths and weaknesses of individual students for both the reading selection genres and the skill sets for each ladder.</td>
<td></td>
</tr>
<tr>
<td>• If necessary or desired, re-group students based on strengths and weaknesses that are apparent from the assessment sheets.</td>
<td></td>
</tr>
<tr>
<td>• Assign students additional Jacob's Ladder readings chosen specifically for them based on their strengths and weaknesses. For example, if the assessment sheets indicate a student's weakness with Ladder A, choose reading selections that have a corresponding Ladder A for the student to complete. Also, if a</td>
<td></td>
</tr>
</tbody>
</table>
| Week Two (con’t) | student shows a particular strength in Ladder C, you may wish to assign reading selections with a corresponding Ladder C to provide students with an opportunity to experience success with a new curriculum and new thinking strategy.  
- All students do not need to be assigned the same reading selection.  
- Students should complete 3 *Jacob’s Ladder* reading selections during Week Two following the procedures as outlined in Week One. |
| Week Three (Observations will occur during this week) | • Continue monitoring student assessment sheets and re-grouping and/or assigning readings accordingly.  
- Students should complete 2 to 3 *Jacob’s Ladder* reading selections during Week Three. |
| Week Four | • Continue monitoring student assessment sheets and re-grouping and/or assigning readings accordingly.  
- Students should complete 2 to 3 *Jacob’s Ladder* reading selections during Week Four. |
| Week Five | • Continue monitoring student assessment sheets and re-grouping and/or assigning readings accordingly.  
- Students should complete 2 to 3 *Jacob’s Ladder* reading selections during Week Five. |
| Week Six (Observations will occur during this week) | • Continue monitoring student assessment sheets and re-grouping and/or assigning readings accordingly.  
- Students should complete 2 to 3 *Jacob’s Ladder* reading selections during Week Six. |
| Week Seven | • Continue monitoring student assessment sheets and re-grouping
### PILOT STUDY OF JACOB'S LADDER

<table>
<thead>
<tr>
<th>Week</th>
<th>Instructions</th>
</tr>
</thead>
</table>
| Week Eight | - This is the last week of implementation.  
- Continue monitoring student assessment sheets and re-grouping and/or assigning readings accordingly.  
- Students should complete the last reading selection in the first two genre sections of the *Jacob's Ladder* notebook during Week Eight plus an additional ladder of teacher's choice. |
| Week Nine | - Administer the TCT following the instructions provided in the manual.  
- Administer the Reading subtest of the ITBS following the instructions provided in the manual.  
- Complete the teacher feedback form and ask students to complete the student feedback form.  
- Return all assessments and feedback forms to the researcher. |
PILOT STUDY OF JACOB’S LADDER

APPENDIX H:

Emergent Themes and Representative Comments from Teacher Feedback Form
## PILOT STUDY OF JACOB'S LADDER

<table>
<thead>
<tr>
<th>Emergent Themes</th>
<th>Representative Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>STUDENT RESPONSE</strong></td>
<td></td>
</tr>
<tr>
<td>Student enjoyment</td>
<td>• Students loved it</td>
</tr>
<tr>
<td></td>
<td>• Students loved the poetry</td>
</tr>
<tr>
<td></td>
<td>o would ask, “Can we read a poem?”</td>
</tr>
<tr>
<td></td>
<td>• Loved the word “haunches” in the Fog poem—used the word for weeks after reading the poem</td>
</tr>
<tr>
<td></td>
<td>• No resistance from students</td>
</tr>
<tr>
<td></td>
<td>• Students liked the choice and the variety of materials</td>
</tr>
<tr>
<td></td>
<td>• One student created a poetry book over the weekend because of work with curriculum; on her acknowledgements page she thanked her teacher for giving her the opportunity to experience poetry in this way</td>
</tr>
<tr>
<td></td>
<td>• Students wanted to read and take notes</td>
</tr>
<tr>
<td></td>
<td>• Students were telling other teachers about the work they were doing with Jacob's Ladder</td>
</tr>
<tr>
<td></td>
<td>• Students enjoyed working together</td>
</tr>
<tr>
<td>Student discussions</td>
<td>• Students loved the discussions because it was the first time they could discuss literature on this level with each other</td>
</tr>
<tr>
<td></td>
<td>• Teachers loved listening to the students discussions</td>
</tr>
<tr>
<td></td>
<td>• Students supported each other during discussions</td>
</tr>
<tr>
<td></td>
<td>• Liked using the curriculum with reading groups so the teacher could listen in on student discussion</td>
</tr>
<tr>
<td>Positive response to fiction story length</td>
<td>• Students were surprised and pleased by the shortness of the stories</td>
</tr>
<tr>
<td>Disliked the nonfiction</td>
<td>• Students did not engage in discussions of nonfiction—it did not spark their interest</td>
</tr>
<tr>
<td></td>
<td>• Nonfiction was horrible</td>
</tr>
<tr>
<td></td>
<td>• too overwhelming for students</td>
</tr>
<tr>
<td>Could relate to the subject matter</td>
<td>• Students could relate to the poems especially, in particular, “My Sister is a Sissy” and “Cousin for Sale”</td>
</tr>
<tr>
<td>Emergent Themes</td>
<td>Representative Comments</td>
</tr>
<tr>
<td>-----------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Students surprised teacher with their ability to grasp abstract concepts | • One teacher noted that the students who never completed their work, never finished a book, and never turned anything in were the students who did best with this curriculum in her classroom  
• A 3rd grade student discovered that her hidden talent was critical thinking as a result of using this curriculum  
• The 5th grade teacher was surprised when her students really “got” Frost’s poem *The Road Not Taken*  
• One teacher noticed a low-reading student really engaged in the curriculum and actively choosing to work with higher ability students as a way to improve his own work |
| **INSTRUCTION**                                     |                                                                                                                                                                                                                           |
| Students had difficulty with generalizations, inferences, and synthesizing | • They asked the most questions about generalizing, synthesizing, and drawing inferences  
• Students had particular difficulty formulating their own generalizations  
• Students asked to look at the Athena generalizations poster  
• Teacher manual doesn’t give clear description of generalizations with examples  
• Manual should detail the Taba Model process with examples  
• One teacher went through the Taba Model as a whole class several times; then had students complete first two levels independently while continuing to create generalizations as a whole class |
| Reading level was too difficult (5th grade)         | • Students had difficulty with the reading level of some selections, e.g. *The Gettysburg Address*                                                                                                                       |
| Ladder terminology requires direct teaching         | • Had to teach a lot of the ladder words, e.g. generalization and inferences  
• Needs to be taught before exposing students to the ladders  
• Needs to be taught in the context of each individual ladder |
## PILOT STUDY OF JACOB'S LADDER

### Emergent Themes

<table>
<thead>
<tr>
<th><strong>IMPLEMENTATION</strong></th>
<th><strong>Representative Comments</strong></th>
</tr>
</thead>
</table>
| Flexibility and organization of curriculum | • Teacher was able to select pieces she thought the students would be able to handle  
| | • Was very simple to implement  
| | • Curriculum is well organized  
| | • The time you spend on assessment is gained back in planning time  
| | • Loved the simplicity  
| | • Teachers and students appreciated the flexibility  
| | • Liked being able to move among genres and ladders—kept it “fresh”  
| | • Students could present their answers in a variety of ways  |
| Assessment system | • You learn a lot about your students while assessing their work  
| | • Didn’t like the assessment system because it was too hard to translate into “real grades”  |
| Time required | • Implementation took longer than expected—anywhere from 2 days to one week per reading selection and corresponding ladder  
| | • Curriculum was not given the time it deserved: “The students wanted to do well. I could see their motivation. We didn’t give it the time it deserves.”  |
| Timing of implementation | • Curriculum was rushed because it was the end of the school year  
| | • Teachers are concerned about the validity of posttest results because  
| | o students were “out of test mode”  
| | o had just completed state assessments  
| | o were distracted by other school activities such as art day in the class next door  
| | o they were given the next to the last week of school  
| | o the amount of time between the pre- and posttests was short  |
PILOT STUDY OF JACOB'S LADDER

<table>
<thead>
<tr>
<th>Emergent Themes</th>
<th>Representative Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDUCATIONAL BENEFIT</td>
<td></td>
</tr>
<tr>
<td>Promoted individual and collaborative thinking</td>
<td>• “the students were thinking together”</td>
</tr>
<tr>
<td></td>
<td>• Could hear students “figuring out” the answers together</td>
</tr>
<tr>
<td></td>
<td>• Students would seek out each other when they need assistance or wanted to discuss the literature</td>
</tr>
<tr>
<td>Promoted higher level thinking and reasoning skills</td>
<td>• Teacher encouraged students to use reasoning skills to figure out ladder terminology</td>
</tr>
<tr>
<td></td>
<td>• Tasks requiring higher order thinking were already in place—all students had to do was access them</td>
</tr>
<tr>
<td></td>
<td>• Many of the skills were familiar to students but they had to go higher with Jacob's Ladder</td>
</tr>
<tr>
<td>Promotes student independence</td>
<td>• Increased student ability to work independently</td>
</tr>
<tr>
<td></td>
<td>• Higher level students wanted to work independently</td>
</tr>
<tr>
<td>Challenges students</td>
<td>• Students said it made them think</td>
</tr>
<tr>
<td></td>
<td>• The students who say they don’t want to use the curriculum again are showing the value of it—they are being challenged!</td>
</tr>
<tr>
<td>FUTURE USE</td>
<td></td>
</tr>
<tr>
<td>Will use the curriculum next year</td>
<td>• Three teachers said they were looking forward to using the curriculum next year</td>
</tr>
<tr>
<td></td>
<td>• Two teachers (who will not be in a classroom next year) said they would recommend the use of the curriculum to their students’ new teachers</td>
</tr>
<tr>
<td>Use for entire school year</td>
<td>• Would integrate with book clubs and use throughout the year</td>
</tr>
<tr>
<td></td>
<td>• In an ideal situation, teacher would use for the entire year</td>
</tr>
<tr>
<td></td>
<td>• Would give the curriculum the time it deserves</td>
</tr>
<tr>
<td>Use beginning 2nd marking period</td>
<td>• Would pre-teach terminology during the first marking period and then begin using the curriculum</td>
</tr>
<tr>
<td></td>
<td>• Using Jacob's Ladder during the first marking</td>
</tr>
</tbody>
</table>
PILOT STUDY OF *JACOB'S LADDER*

<table>
<thead>
<tr>
<th>Emergent Themes</th>
<th>Representative Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CURRICULUM IMPROVEMENTS</strong></td>
<td></td>
</tr>
<tr>
<td>Add more readings/ladders to each genre</td>
<td>• Add more readings and ladders to myths/fables and poetry in particular</td>
</tr>
</tbody>
</table>
| Add more variety to reading levels within each level of curriculum | • There should be a wider range of reading levels within the Level III *Jacob's Ladder* so the students who need more challenge can have it, but there are also selections for students who are not quite ready to read such texts as *The Gettysburg Address*  
  • The reading level of the nonfiction books at 3rd grade is too difficult |
| Improvements to nonfiction                          | • When creating ladders, divide the books into sections with separate ladders for each section  
  • Teachers really need a class set of at least one nonfiction book to facilitate modeling with the whole class  
  o Should package as 25 copies of one book with 5 copies of each additional title |
| Answer sheet                                         | • The answer sheet was too confining and too limiting  
  • The answer sheet needs lines  
  • The answer sheet should be used only at the beginning as a model of how students should organize their answers and assess their own work |
APPENDIX I:

Student Responses to Questions 1-4 on Feedback Form
## Response to Student Feedback Question 1

<table>
<thead>
<tr>
<th></th>
<th>YES</th>
<th>Rationale</th>
<th>NO</th>
<th>Rationale</th>
<th>UNDECIDED</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Rationale</td>
<td>N</td>
<td>Rationale</td>
<td>N</td>
<td>Rationale</td>
</tr>
<tr>
<td>Question 1:</td>
<td>59</td>
<td>- the readings were funny</td>
<td>13</td>
<td>- because I don't like reading</td>
<td>3</td>
<td>- I liked the nonfiction books best</td>
</tr>
<tr>
<td>Enjoyment of</td>
<td></td>
<td>- it can help kids like reading</td>
<td></td>
<td>- because the cause and effects are hard</td>
<td></td>
<td>- some were boring; they didn't catch my interest or entertain me</td>
</tr>
<tr>
<td>readings in the program</td>
<td></td>
<td>- the stories were interesting</td>
<td></td>
<td>- because it was hard</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- they taught you a lesson</td>
<td></td>
<td>- they could be more detailed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- I learned wonderful things from them</td>
<td></td>
<td>- there was less action than I expected</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- they were fun</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- I love to read</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- I just love poetry</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- the stories were good</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- they were enjoyable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- most of them were exciting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- there were science books</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- they were empowering and made me think</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- because they were harder than Trophies books</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- you can challenge yourself</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- because the stories were well written</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**PILOT STUDY OF JACOB'S LADDER**

**Response to Student Feedback Question 2**

<table>
<thead>
<tr>
<th>YES</th>
<th>NO</th>
<th>UNDECIDED</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>Rationale</td>
<td>N</td>
</tr>
</tbody>
</table>
| 43  | - we got to speak our minds and tell other people what we think  
- you could see if you were right  
- you are listening  
- you learn what mistakes you made  
- you get to talk with a friend  
- I like to share my ideas  
- I got to hear what other people had to say about the selection  
- I liked explaining the easy questions  
- it helped me understand some of the questions  
- it's more challenging than just answering them  
- it gave me more experience with bigger, harder words  
- a lot of the discussions made me think  
- I like getting into detail  
- it made the questions easier | 31  | - I thought some people had wrong answers  
- it always ended in a mess  
- it's sometimes hard to answer questions and I often don't understand them  
- because I don't get along with some students  
- everybody had their own point of view so it got really confusing  
- I felt that some people wouldn't understand my reasons  
- because I didn't learn anything  
- I had trouble understanding the questions  
- because it took a long time  
- because I might have funny answers  
- I'm shy  
- sometimes I put the wrong answer  
- I'm afraid of getting it wrong and being embarrassed  
- most of the time it was boring  
- I would rather work on it alone  
- I don't like to share my answers | 1   | - some questions I didn't get |
### Response to Student Feedback Question 3

<table>
<thead>
<tr>
<th>N</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES</td>
<td>I learned what you have is good enough</td>
</tr>
<tr>
<td>NO</td>
<td></td>
</tr>
</tbody>
</table>

- Never give up
- Learned about topics, e.g., animals, martyrs, storms, bugs, science, space, etc.
- Make sure you read every question
- Go back to the story
- Try
- Things may not be what they seem
- Lessons from the fables
- Read carefully
- Better vocabulary
- How to write poems
- Better reading skills
- What generalizations are
- Thinking harder
- How to fill out a ladder
- How to pay close attention to questions
- Confidence
- Respect
- Compassion
- Read between the lines
- Reading helps you learn and discover
- What an inference is
- What a consequence is
- Sequencing/ordering
- How to cooperate with others
- How to understand poems more
- How to support my answers with the reading
- How to take better notes
- Grading myself fairly

### Response to Student Feedback Question 4

<table>
<thead>
<tr>
<th>N</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES</td>
<td>I learned a lot</td>
</tr>
<tr>
<td>NO</td>
<td>Because they are difficult to do</td>
</tr>
</tbody>
</table>

- It is helpful
- It is fun to do
- The stories are interesting
- It will make me smarter
- I would like to get better at looking back for the answers
- It will be more challenge for me
- It was hard enough this year
- I don't like to read
- It is hard to understand
- It takes too long
- We never got to pick our own stories
- It wasn't as fun as I thought it would be
- It was hard enough this year
- I don't like to read
- It is hard to understand
- The questions were frustrating
- It takes too long
- We never got to pick our own stories
- I don't like to discuss questions
- It wasn't as fun as I thought it would be

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