

MAKING STUDENTS' WRITING BLOOM: THE EFFECT OF SCAFFOLDING  
ORAL INQUIRY USING BLOOM'S TAXONOMY ON WRITING IN  
RESPONSE TO READING AND READING COMPREHENSION  
OF FIFTH GRADERS

Except where reference is made to the work of others, the work described in this dissertation is my own work or was done in collaboration with the advisory committee. This dissertation does not include proprietary or classified information.

---

Brooke Allen Anthony

Certificate of Approval

---

Edna G. Brabham  
Associate Professor  
Curriculum and Teaching

---

Bruce A. Murray, Chair  
Associate Professor  
Curriculum and Teaching

---

Alyson I. Whyte  
Associate Professor  
Curriculum and Teaching

---

George T. Flowers  
Interim Dean  
Graduate School

MAKING STUDENTS' WRITING BLOOM: THE EFFECT OF SCAFFOLDING  
ORAL INQUIRY USING BLOOM'S TAXONOMY ON WRITING IN  
RESPONSE TO READING AND READING COMPREHENSION  
OF FIFTH GRADERS

Brooke Allen Anthony

A Dissertation

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements of the Degree of

Doctor of Philosophy

Auburn, Alabama  
May 10, 2007

MAKING STUDENTS' WRITING BLOOM: THE EFFECT OF SCAFFOLDING  
ORAL INQUIRY USING BLOOM'S TAXONOMY ON WRITING IN  
RESPONSE TO READING AND READING COMPREHENSION  
OF FIFTH GRADERS

Brooke Allen Anthony

Permission is granted to Auburn University to make copies of this dissertation at its discretion, upon request of individuals or institutions and at their expense. The author reserves all publication rights.

---

Signature of Author

---

Date of Graduation

## VITA

Brooke Allen Anthony, daughter of Marvin Cecil and Jacqueline Allen Anthony, was born on February 4, 1972, in Silver Spring, Maryland. She graduated from Hood College with a Bachelor of Arts degree in Early Childhood Education, earned a Master of Science degree in Reading Education from Johns Hopkins University in 2000, and earned a Specialist in Education degree in Elementary Education from Auburn University in 2003. Throughout her graduate work, Ms. Anthony had previously been employed full time as an elementary school teacher, reading specialist and currently an Assistant Principal. She is the mother of three children, Rachael Elizabeth, Emily Allen, and Griffin Pate Anthony.

MAKING STUDENTS' WRITING BLOOM: THE EFFECT OF SCAFFOLDING  
ORAL INQUIRY USING BLOOM'S TAXONOMY ON WRITING IN  
RESPONSE TO READING AND READING COMPREHENSION  
OF FIFTH GRADERS

Brooke Allen Anthony

Doctor of Philosophy, May 10, 2007  
(Ed.S., Auburn University, 2003)  
(M.S., Johns Hopkins University, 2000)  
(B.A., Hood College, 1997)

100 Typed Pages

Directed by Bruce A. Murray

This pretest-posttest control group study attempted to investigate the effects of using Bloom's Taxonomy as an oral-questioning scaffold to improve writing in response to reading and reading comprehension by encouraging higher order thinking. Participants, 22 fifth-grade students from a suburban school, were randomly assigned to control and experimental groups.

Writing was assessed in two ways; using a holistic score and a point system. The testing instrument was a researcher-created rubric based on Bloom's taxonomy.

Repeated-measures ANOVA revealed that the use of higher order questioning improves writing in response to reading, whether scored holistically or with points. Writing scores showed high interrater reliability. The study showed that when teachers instruct students using a higher order questioning scaffold based on Bloom's taxonomy, writing significantly improves.

The study also investigated the effect of higher order questioning on reading comprehension. Reading comprehension was assessed in two ways; using the Degrees of Reading Power (DRP), a standardized test of reading comprehension, and using a researcher-created test. Neither test showed a significant advantage for higher level questioning over lower level questioning on reading comprehension, although on the DRP standardized assessment, the treatment group showed a strong trend toward greater achievement ( $p = .06$ ). This study provided preliminary support to the importance of using higher order thinking as a questioning scaffold.

## ACKNOWLEDGEMENTS

I would like to express my genuine appreciation to all who have provided support and encouragement throughout this study, with special recognition given to Dr. Bruce Murray, chair of my committee, without whose guidance this study would have never been finished. I also wish to express my great appreciation to the other committee members, Dr. Edna Brabham and Dr. Alyson Whyte, for their expertise and knowledge throughout my doctorate program. Additionally, I appreciate the countless hours Jim Devilbiss spent helping me with my research data and statistical analysis. I further would like to thank my parents, Marvin Cecil and Jacqueline Allen Anthony, and sisters Leslie Anthony Mangold and Carrie Anthony Hines for their endless support, encouragement and for the high expectations they taught me to have in myself. In a final note, I wish to extend my deepest gratitude, commitment and love to my three children: daughter, Rachael Elizabeth, daughter Emily Allen, and son, Griffin Pate Anthony. To all three children I dedicate this work in hopes that Rachael, Emily and Griffin will strive to achieve what may seem unattainable while stopping to smell the roses along the way.

Style manual used: Publication of the American Psychological Association (Fifth Edition)

Computer software used: Microsoft Word, Minitab, SPSS

Books used for weekly instruction:

*Bread and Jam for Frances* by Russell Hoban

*Ms. Nelson is Missing* by James Marshall

*Stone Soup* by Ann McGovern

*The Paper Crane* by Molly Bang

## TABLE OF CONTENTS

	Page
LIST OF TABLES.....	xi
CHAPTER I. INTRODUCTION.....	1
Introduction.....	1
History.....	2
Theoretical Framework.....	4
Significance of the Study.....	6
Research Questions.....	7
Preview.....	8
Definition of Terms.....	10
CHAPTER II. REVIEW OF LITERATURE.....	12
Teacher and Student Questioning – Guiding Oral Inquiry.....	12
Questioning – Scaffolding as a Strategy.....	14
Assessing Writing.....	17
Assessing Higher Order Thinking.....	26
Accountability.....	27
CHAPTER III. METHODOLOGY.....	29
Research Question.....	29
Setting.....	30
Pilot Testing.....	30
Participants.....	31
Outcome Variables.....	32
Research Design.....	34
Analysis of Data.....	38
Projected Qualitative Analysis of Data.....	39
CHAPTER IV. DATA ANALYSIS AND RESULTS.....	40
Study Overview.....	40
Findings Related to Question 1.....	41

Findings Related to Question 2.....	44
Findings Related to Question 3.....	46
Findings Related to Question 4.....	47
Findings Related to Question 5.....	50
Findings Related to Question 6.....	54
Summary.....	57
 CHAPTER V DISCUSSION.....	 59
Conclusion.....	59
Interpreting Data.....	59
Contributions to Literature.....	70
Practical Implications of the Study.....	71
Limitations.....	73
Recommendations for Future Research.....	73
 WORKS CITED.....	 77
 APPENDIX A Higher Order Thinking Rubric.....	 80
 APPENDIX B <i><u>Lucky Cricket</u></i> – Pretest for Reading Comprehension.....	 82
 APPENDIX C <i><u>Father’s New Game</u></i> – Posttest for Reading Comprehension.....	 84
 APPENDIX D Student Assent Form.....	 86
 APPENDIX E Informed Consent Form.....	 87

## LIST OF TABLES

1. Paired T-Test for the Experimental Group Holistic Scores Pretest and Posttest.....	42
2. Paired T-Test for the Experimental Group Point Scores Pretest and Posttest.....	43
3. Paired T-Test for the Experimental Group DRP Test of Reading Comprehension Pretest and Posttest.....	44
4. Paired T-Test for the Experimental Group Researcher Test of Reading Comprehension Pretest and Posttest.....	45
5. Paired T-Test for the Control Group Holistic Scores Pretest and Posttest.....	46
6. Paired T-Test for the Control Group Point Scores Pretest and Posttest.....	47
7. Paired T-Test for the Control Group DRP Test of Reading Comprehension Pretest and Posttest.....	48
8. Paired T-Test for the Control Group Researcher Test of Reading Comprehension Pretest and Posttest.....	49
9. Descriptive Statistics for the Between Subjects Pretest Posttest Holistic Scores of the Control and Experimental Groups.....	50
10. Repeated Measure Analysis of Variance for the Pretest Posttest Holistic Scores of the Control and Experimental Groups.....	51
11. Descriptive Statistics for the Between Subjects Pretest Posttest Point Scores of the Control and Experimental Groups.....	52
12. Repeated Measure Analysis of Variance for the Pretest Posttest Point Scores of the Control and Experimental Groups.....	52
13. Descriptive Statistics for the Between Subjects Pretest Posttest Scores on the DRP test for the Control and Experimental Groups.....	55
14. Repeated Measure Analysis of Variance for the Pretest Posttest DRP Scores of	

the Control and Experimental Groups.....	55
15. Descriptive Statistics for the Between Subjects Pretest Posttest Scores on the DRP test for the Control and Experimental Groups.....	56
16. Repeated Measure Analysis of Variance for the Pretest Posttest Researcher- created Test Scores of the Control and Experimental Groups.....	57

## CHAPTER I

### INTRODUCTION

On January 8, 2002, President George W. Bush signed the new education law “No Child Left Behind.” This act encompasses four basic principles for educational reform: an emphasis on teaching methods that have been proven to work through scientifically based research, expanded options for parents, increased local control and flexibility, and stronger accountability for results. This law is the largest reform act of its kind since the Elementary and Secondary Education Act (ESEA) of 1965. While the four principles are not new ideas in the field of education, they have left states and local school systems scrambling to meet the strict guidelines.

Currently Maryland and other states are now requiring proficiency on state exams in order for students to graduate high school. As a 10 year educator in Maryland, this impacts my students greatly. The Maryland State Assessment (MSA) requires students to show their understanding of text, through writing on the MSA. While the theory of central tendency claims that not all students will achieve the same, No Child Left Behind (NCLB) says that all students can achieve average or higher than average scores. While theoretically all educators agree in raising the standard and setting the bar high, classroom teachers and teachers of students with special needs would argue that some students may never reach this goal. It’s the parents of these students, who beg educators to tell them how to help improve their child’s reading and writing so they can pass the rigorous state testing and graduate high school.

This study addressed two of the four principles of NCLB, accountability through performance assessment in writing, and using scientific based research to improve instructional strategies. Bloom's Taxonomy was the questioning scaffold used to provide higher order question stems in hopes that the deeper comprehension accessed by participants would improve overall writing in response to reading and reading comprehension.

Perhaps the results of this study will help parents and educators alike, find a way to improve students' reading comprehension and written response, so that all students can learn to think more critically, earn a high school diploma and not be a child "left behind."

### History

In recent years a newer kind of testing, performance based assessment, has been linked to educational reform hence the great debate over performance based assessment verses standardized assessment has received much attention. Performance based assessment promotes hands-on, experiential activities that are more closely aligned with best practice (Seal, 1993). Standardized assessment consists of traditional types of questioning using true-false, matching and multiple choice varieties and may typically require one right answer (Strange, 1997). It is important to consider the use of performance-based assessment because one dependent variable in this study is the measurement of higher order thinking seen in students' writing. There is no shortage of opinion on either testing format.

Performance based assessment has many advocates. It is considered a closer measure of a child's ability not only in the ability to retain and recall information but also

in the demonstration of knowledge. One author suggests, “If you want to see if someone can ride a bike, you don’t give them a multiple choice test. You see if he can do it (Seal, 1993).” The word “performance” in itself requires an active measure. This belief stems as far back as the early 1900s when one of the most influential educators, John Dewey revealed his theory on learning. Dewey believed that learning is an active involvement starting with assimilation within (Dewey, 1902). In essence, learning and thinking is what an individual does with what he studies. Zemelman, Daniels and Hyde (1998), suggests that assessment, in its best form, is authentic. It isn’t filling in the blank papers and simply a regurgitation of learned. Authentic assessment is where children re-create and reinvent what they encounter (Zemelman, et al., 1998). Mitchell agrees stating that performance assessment is active student production as evidence of learning. It is not a collection of responses that entail a passive selection of the one right answer in the context of preconstructed answers (Mitchell, 1995). Using this type of assessment, students develop higher order thinking not just the ability to calculate and memorize (Seal, 1993). As stated in Goals of America 2000, if students are to “use their minds well” then new assessment techniques need to be established (Strange, 1997). In contrast, critics argue that scoring performance-based tests can be subjective and expensive. Nevertheless many states have now adopted performance-based testing measures as a substitute for standardized assessment or for use concomitantly. While many researchers firmly believe in this kind of measure, many feel differently.

No Child Left Behind has rejuvenated proponents of standardized tests. “We’re results-oriented people in this country,” President Bush commented. “In return for taxpayers’ money, we ought to insist upon results (Borsuck, 2001).” The testing

measure however is undisclosed and probably left up to the local districts. Mehrens suggests that performance assessment is but a window dressing and the real stuff is in the standardized form of multiple-choice format (Mehrens, 1992). Strange (1997) comments that those who think multiple-choice tests don't measure higher order thinking are not thinking high level enough (Strange, 1997). The literature review will compare and contrast the application of higher order thinking more thoroughly.

What are higher order, otherwise known as critical, thinking skills? Higher order thinking, involves many things to which there is no set definition. Basically, higher order thinking involves different types of thinking. Individuals often need to see relationships among things and move beyond the literal to include divergent, convergent, inductive, deductive, open-ended questions and creative thinking. One of the most influential educational psychologists of his time, Benjamin Bloom, created a six level classification system to offer educators a design or scaffold as it will be referred to, that encompasses critical thinking, higher order questioning and constructivism. This system became known as Bloom's taxonomy (Bloom, Engelhart, Furst, Hill, Krathwohl, 1956).

### Theoretical Framework

Benjamin Bloom was an educational psychologist who believed that all learners can succeed (Bloom, 1956). With the help of his colleagues, in 1956 he developed this system to help graduate students in their writing of literature reviews. It was intended as a "means of facilitating communication and improving the exchange of ideas (Benson, 1992)." Although solely Benjamin Bloom did not create Bloom's *Taxonomy of Educational Objectives*, it became known as "Bloom's taxonomy."

Bloom's taxonomy consists of six levels, knowledge, comprehension, application, analysis, synthesis and evaluation. Knowledge is the lowest level and evaluation is the highest level. The leveling system is a hierarchy of words and concepts used as questioning stems or the command words used at the beginning of questions. Knowledge words are list, tell, who, what, when, where, etc... Evaluation words are justify, validate, critique, debate, etc... The focus of using Bloom's taxonomy in education aims to shift the thought processes of students from information gathering to information processing. This requires active participation not passive recollection, a constructivist approach.

Bloom's taxonomy was initially created to assist students in counselor education (Granello, 2000), more recently it has been used to promote higher level, critical thinking in the elementary classroom. Bloom's Taxonomy has been used in numerous ways from the classroom to the corporate world. Application of this model crosses all content areas and can be used with any age student. Bloom's taxonomy is very versatile and has been widely accepted as a way to promote higher order thinking. Using Bloom's taxonomy as a scaffold, teachers can assess reading comprehension by facilitating conversation and the exchange of ideas (Granello, 2000). As teachers use this scaffold, the teacher is able to identify the level of reading comprehension while sharpening and clarifying the way students think critically. Teachers can identify the level of the comprehension, and then challenge students using an oral inquiry scaffold of higher level questioning based on Bloom's Taxonomy. "Research indicates that the level of the taxonomy that the teacher uses influences the level of response among students (Wilson, 1973)."

In this study, Bloom's taxonomy provided a scaffold for critical thinking. I believe that the use of this questioning scaffold in classrooms will increase students'

higher order thinking as shown in their writing and will improve reading comprehension as a result of this treatment. As stated, this study used oral inquiry as a questioning scaffold, it is important to discuss oral inquiry.

Oral inquiry is another way of questioning students aloud. Through oral inquiry, conversations about text perpetuate and scaffold students toward higher level, critical thinking. As students gain experience and become comfortable discussing text this way, they will begin applying these skills in their writing. “Through writing, the teacher can identify the cognitive level and style of the writer (Granello, 2000).” As with reading comprehension, once the teacher identifies the taxonomic level of the students’ writing, he/she can work with the child towards achieving the next level. Several theoretical components may call into question the results from this study. Readers not in agreement with constructivism as a hands-on activity and who envision the meaning of this theory differently doubt the theoretical analysis behind this perspective. Furthermore, there may be doubts as to whether the writing is considered a constructivist response. Colleagues may not believe that using Bloom’s taxonomy as a questioning scaffold is a method of improving higher order thinking. Finally, opponents may not agree that holistic writing is a true measure of quality writing.

### Significance of the Study

This study contributed to a greater understanding of how the use of higher order thinking affected student achievement in reading comprehension and writing. Previously, limited research on higher order thinking provided statistical data to support the general theory that teachers should be using higher order thinking in the classroom. As each

school year begins, teachers across the nation are given handouts, posters, attend workshops and professional development sessions advocating the use of higher order thinking in the classroom, but little research to date has shown teachers why this method of teaching is important and practical.

Furthermore, as previously mentioned Maryland and many other states are raising educational standards for students and requiring students to pass rigorous state testing yearly in preparation for a high school diploma down the road. Additionally we require teachers to teach using research based practices but have failed to provide research to support the importance of instruction using higher order thinking.

The data from this study offers teachers, administrators, curriculum specialists and all stakeholders invested in the education of children, true experimental research to support a common belief that higher order thinking is important.

Lastly, as educators our larger goal is to provide tools to enable students to be successful today in classrooms and later in life. If we want to help students unlock the power of experiencing success in reading and writing, then higher order thinking is the key.

#### Research Question

The purpose of this study was to investigate the effects using Bloom's Taxonomy as an oral questioning scaffold to improve writing in response to reading and reading comprehension through the use of higher order thinking.

Students had the opportunity to greatly benefit from this study. Through exposure to a hierarchy scaffold of critical thinking questions, some students may have learned

how to think more critically in class and independently. Perhaps higher order questions enabled students to organize and rehearse thoughts better. Hence, students' writing improved. Other students had the opportunity to use visualization as an instructional strategy to better comprehend text and improve writing.

The results of this study showed educators how using a scaffold of oral inquiry produces readers who independently think critically, display higher order thinking in writing and have improved reading comprehension. Perhaps the data from this research will encourage teachers to try using this scaffold based on Bloom's taxonomy.

#### Preview

While many educators question the validity of performance-based assessment, most seem to agree in the benefits of learning higher order thinking. Benjamin Bloom has played a major role in developing a strategy to reinforce and further the use of critical thinking in the classroom. This study investigated the effects of using Bloom's Taxonomy as an oral questioning scaffold to improve writing in response to reading and reading comprehension through the use of higher order thinking. The research questions to be investigated are:

1. Do students demonstrate higher level thinking in their writing in response to teacher questioning from higher levels of Bloom's taxonomy?
2. Do students demonstrate better reading comprehension in response to teacher questioning from higher levels of Bloom's taxonomy?
3. Do students demonstrate higher level thinking in their writing in response to teacher questioning from lower levels of Bloom's taxonomy?

4. Do students demonstrate better reading comprehension in response to teacher questioning from lower levels of Bloom's taxonomy?
5. Does teacher questioning from higher levels of Bloom's taxonomy cause students to demonstrate higher level thinking in writing than does teacher questioning from lower levels of Bloom's taxonomy?
6. Does teacher questioning from higher levels of Bloom's taxonomy improve reading comprehension more than teacher questioning from lower levels of Bloom's taxonomy?

## Definition of Terms

**Standardized Assessment** – True/False and Multiple Choice type assessment where there is only one correct answer

**Performance Based Assessment** – Assessment where students are tasked to create a response. There is no one correct answer and students can meet the criteria set forth in many different ways.

**Rubric** – A hierarchy assessment tool, typically 3 - 6 levels, where the overall work is scored according to set criteria and/or anchor paper samples

**Holistic Scoring** – Work graded according to rubric specifications

**Scaffold** - A framework for teaching where the teacher uses the knowledge that a student demonstrates, then prompts the student to provide a more in depth, comprehensive and elaborative response, prompting continues in an effort to move the student to higher levels of performance

**MSA** – Maryland School Assessment

**Oral Inquiry** – Teacher questioning of students, typically aloud

**Open Ended Questions** - Questions where the responder answers the question in their own words offering their own knowledge and verbiage in contrast to questions where the correct answer is a predetermined response

**Writing Content** - The meaningful content in writing excluding grammar, punctuation and writing mechanics, often the ideas, elaboration, and explanation in response to the written prompt and purpose

**Questioning Stems** - The beginning question words and commands, for instance, who, what, when, justify, defend, evaluate, etc...

**DRP** – Degrees of Reading Power test for reading comprehension

## CHAPTER II

### REVIEW OF LITERATURE

In order to provide a background of previous research on teacher questioning (guiding oral inquiry), scaffolding as a strategy, writing assessment, assessing higher order thinking skills and accountability to instruction, a thorough search of available research was completed. The goal was to find articles that 1) were published in peer reviewed journals, 2) followed an experimental research design, and 3) provided information to the reader, in contrast to or to corroborate topics pertaining to this study. Although 3 databases were used in the search, few articles on these topics were experimentally designed studies. Therefore, quasi-experimental and qualitative studies were also included in this chapter in an effort to provide the reader with greater background knowledge prior to reading my research.

#### Teacher and Student Questioning – Guiding Oral Inquiry

Teacher questioning of students during reading instruction, often referred to as oral inquiry, has been a topic of interest for many decades. While this has been the traditional method of informally assessing students' reading comprehension, the impact of teacher guided questioning on students' reading achievement is worth investigating. Guszak (1967) conducted a study of 12 teachers (and their classes) to investigate reading and thinking skill development in an elementary reading group. Four teachers at each

levels, second, fourth, and sixth grade, were randomly selected from a population of 106 teachers. This qualitative study involved observing all reading groups in the 12 classrooms for 3 days. Tape recordings of the observations were transcribed and analyzed numerically. The results found that teachers tend to ask mostly literal, recall type questions with little time given to evaluative questioning. Of the more difficult questions asked, majority required simple yes or no answers not allowing students the opportunity to elaborate and explain. Guszak (1967) further found that students answered questions more accurately in the second grade than older grades (mostly likely because the questions in second grade tend to be more factual having only one correct answer). As the grade level and difficulty of text increases, factual recall becomes more difficult because the texts often become dense with details. More complex text also allows for answers that aren't always predetermined with only one correct answer. Lastly, teachers were found to ask follow up questions most likely related to setting and purpose, followed by questions that required students to verify information in the text and ask question types to justify a position. Lastly, making a judgment about the text was observed the least in all grade levels. The implications of this study challenge educators to devote less time to factually driven, literal, recall type questions and encourage classroom teachers to devote more time to asking students why type questions with support and justification of their answer.

In contrast to teacher-generated questions, Davey and McBride (1986) investigated the effects of student generated questioning on reading comprehension, questioning performance, quality of generated questions and accuracy of reading comprehension. This experimental study was a posttest only design with multiple

groups. The sixth grade participants were randomly divided into six groups, five experimental and one control. The five experimental groups were question training, question generation practice, inference question practice, and literal question practice. After the numerous training and practice sessions, students wrote questions about a randomly selected text. The findings show that overall there were positive effects from the question generation group on their generation of questions, accuracy of comprehension and accuracy between predicted versus actual performance. Davey and McBride suggest that teachers need to train students to use this questioning technique to ensure their understanding of literal and inferential meaning in text. Furthermore, this study shows that using the question generation technique improves comprehension. Future research needs to be conducted to determine if these results will remain strong over time. This study statistically shows that students who create their own questions about specific text better comprehend the text.

#### Questioning – Scaffolding as a Strategy

As seen in Guszak's (1967) study, while oral inquiry alone does not lead students to a more in depth understanding of the complexities of text. Wharton-McDonald, R., Pressley, M. and Hampston, J.M. (1998) conducted a qualitative study of 9 first grade teachers to determine what practices and beliefs are present in high quality teachers of reading and writing. The teachers were nominated as outstanding teachers by the curriculum coordinators because of "their observed teaching enthusiasm, reading achievement of students at the year end, writing achievement of students at the year end, teacher involvement in improving their own practice, desire to have their own or their

supervisor's child in their classroom, the teachers abilities with a wide range of abilities and backgrounds and positive feedback" (Wharton-McDonald, 1998). The average teacher in the study had been teaching for 12 years with a range of two to twenty-five years of experience. Of the 9 teachers nominated, 3 emerged as outstanding teachers and were intensely followed, observed and interviewed for approximately 6 months. The results found that highly successful teachers of reading and writing had "coherent and thorough skills with high quality reading and writing experiences, high density of teaching, integrated reading activities, high expectations of all students, masterful classroom management, and an extensive use of scaffolding" (Wharton-McDonald, 1998).

In this study, scaffolding was in the form of questioning, which allowed teachers to monitor responses to text and help students facilitate deeper learning. Three of the 9 teachers, all who were skilled in scaffolding, varied from the intended lesson plan more frequently than teachers whom were less proficient in scaffolding. Teachers were better able to anticipate problems in comprehension and clarify misconceptions during a scaffolding model. Lastly, scaffolding provided teachers more opportunities for on the spot learning and teachable moments.

In contrast to the Wharton-McDonald, et al. (1998) study, question prompts and scaffolding are not only characteristics of outstanding teachers, but scaffolding was found to help students regardless of the facilitator. Ge, X. and Land, S. (2001) study examined the use of question prompts and peer interactions as a scaffolding strategy. This mixed method study was both experimental and quasi experimental. The participants were 115 undergraduate students whose problem solving outcomes, based on a given task, were

measured as a result of the four treatments, individuals with question prompts, individuals without question prompts, peers with question prompts, and peers without question prompts. Observation and interviews were the qualitative instruments measured. Each week 3 sessions (2 class and 1 lab) were measured. The results of the study reported “questioning prompts were a superior scaffolding strategy over peer interactions...” (2001). Additional data suggested that questioning strategies further cognitive growth. While peer interactions were found to benefit the questioning strategies, active participation and engagement were necessary. While this study was conducted at the college level, the findings suggest that achievement gains can be seen with any age participant, when scaffolding is present.

While both Wharton-McDonald, et al. (1998) and Ge, X. and Land, S. (2001), discussed the merits of scaffolding both in outstanding teachers and when used with peers, skeptics question whether students transfer what they learned from a scaffolded learning scenario to a new problem? Murphy, N. and Messer, D. (2000) found that students were able to transfer knowledge learned when adults provided scaffolding. This pretest posttest, control group design measured 122 participants. The participants were video taped while the randomized groups worked through one of three conditions, adult scaffolding, group support, and work alone. The participants were pretested to measure how well they balanced on wooden beams. As part of the research, students were asked to balance familiar objects on a balance beam. Each participant received two chances to balance while either an adult provided scaffolding guiding the participant verbally, a group of participants talked with the participant being tested, or the participant worked alone. The posttest results of the study showed that the participants who had an adult

scaffolding them made significant gains. Participants in the other two condition groups did not make significant gains. This study showed that when students have the opportunity to benefit from scaffolding, their learning does not stay isolated to only the task at hand; newly learned knowledge transfers to other areas.

### Assessing Writing

Empowering students by allowing them to play an active role in their education has been a major characteristic in constructivist classrooms. Writing is considered a constructive expression because it requires the student to understand a concept then recreate and express their understanding in a kinesthetic way. Albertson and Billingsley (2001) investigated student writing with the belief that students taught strategy instruction and self-regulation techniques would show improvement in writing. This quasi-experimental, multiple baseline, time series design focused on two participants who previously worked with the researcher in a similar study. In contrast to other studies, this investigation took place at the researcher's office; therefore the researcher thought it was important to use participants already comfortable in that setting. The two participants were also familiar with the CSPACE technique for prewriting. CSPACE is a mnemonic device referring to a planning handout where C = character, S = setting, P = problem or plot, A = action, C = conclusion, and E = emotion. The design involved three phases, 1) maintenance session where subjects created writing topics and used the CSPACE handout to plan, 2) retraining session where subjects who did poorly in the maintenance session were retrained, and 3) an instructional strategy phase where participants met to plan writing goals. In this study all of the work was completed on the computer. Prior to

the study the subjects were defined as proficient in word processing. The computer was used to measure time spent writing, number of words per minute and rate of writing. Other story characteristics such as locale, plot, action, character development, conclusion and imagination were also measured. Overall, all measurements were taken by two or more raters to ensure consistency and reliability. The results suggest that length, fluency, and text production increased as a result of strategy instruction and self-regulation techniques. In both cases, the participants surpassed their writing goals. Correlations of 98% to 99% indicate that self-regulation improves writing performance. This is applicable to the classroom, suggesting that students should play an active part in writing to demonstrate understanding, and in goal setting as a strategy to improve writing.

In the Albertson and Billingsley study (2001), the reader was left wondering what type of strategy instruction, improves writing ability? Future research needs to be conducted to see if the results are reliable. With a sample this small, it was unclear whether these results would be consistent with a larger sample. In addition, this study used two gifted students that were identified as quick writers but produced writing below potential. Experimenter bias may have also contributed to the validity of the findings in this study whereas this researcher had worked previously with these students. All in all, while proponents believe that having students take a vested interest in their learning promotes better life long learners, because of the small sample sized used Albertson and Billingsley were unable to provide significant data to support this claim.

Similarly while Albertson and Billingsley (2001) had students set writing goals for themselves, Ross J., Rolheiser, C., and Hogaboarn-Gray, A. (1998) had students rate themselves to promote better student writing. The purpose of their study was to

investigate whether self-evaluation training increased the accuracy of students' self-assessment and will this training contribute to language achievement. This quasi-experimental study involved teachers and their students from two different school districts and focused solely on narrative writing tasks. While there was a control group, initially, teachers in the school districts volunteered themselves and students to participate in this study. Fifteen classes in one district were matched to fifteen classes in another with one set representing the treatment group and the other set representing the control group. There was no indication that the groups were equally matched. After this division, some randomization of students selected for analysis occurred. The design of this study involved conducting surveys, writing short stories and evaluating stories that were written. The teachers were trained after school prior to demonstrating the self-evaluation techniques throughout twelve practice sessions. From the treatment group, pre and posttest writing assignments were collected and scored according to a writing rubric. The results indicate that treatment students significantly outperformed control group students on the pre-writing task. The treatment group was also significantly more accurate on the self-assessment. The treatment group had a moderate effect size of 40% overall showing that students who are taught to evaluate their work through the self-evaluation training, wrote better on narrative writing tasks. This study questioned the effects of self-evaluation training on language achievement; however this issue wasn't specifically addressed. The scoring of surveys and the specifics regarding the rubrics were not discussed. Perhaps further research should be conducted on the definition of self-efficacy and the criteria for judging that in oneself. This study also discussed other raters and the scoring criteria used including interrater reliability. The sample study of

296 students seems to be rather large, and I imagine somewhat overwhelming to study. There was no mention of difficulties involved with using a sample this size. I wonder what were the limitations associated with this sample size? Overall this study adds more evidence that authentic assessment of writing, through self-evaluation, more accurately measures student abilities when teachers are trained to teach this strategy. According to this study, participants are able to share the responsibility of assessing their own work and accurately do so. Like Albertson and Billingsley's (2001) study, teachers and students would benefit from being partners in the process of creating writing, not just in obtaining the final product. With regard to my study, participants also play an active role in the comprehension of texts and skills associated with higher order thinking. In contrast however, the subjects' engagement first comes in the form of participation in an oral questioning scaffold with the anticipation (or hope) that subjects will transfer those ideas to writing.

In comparison to Ross et al. (1998) study on the merits of self-evaluation in writing, Pollington, M., Wilcox, B., and T. Morrison (2001) conducted a comparative study on self-perception in writing. In this study participants rated themselves on a specific test design. The purpose of the study was to compare self-perception of writing, among students who were taught through traditional writing instruction and those taught with writer's workshop approach. This experimental design study randomly assigned fourth and fifth grade students to classes prior to the beginning of the school year. Using extreme case selection, eight out of twelve teachers were asked to participate in the study. Four times a year the eight teachers were observed for one hour to ensure consistency in instruction. Teachers were asked if they'd like to teach the experimental (Writer's

Workshop) or the control group (traditional writing instruction). Asking the teachers their preference reduced the likelihood that would be teachers uncomfortable or unfamiliar with writing strategies and methods. Pretesting did not occur. The posttest used was the nationally normed Writer's Self Perception Test (WSPT). The WSPS measures five dimensions: general progress, specific progress, observational comparison, social feedback and physiological states. This instrument has 38 statements about writing. Scores were quantitatively measured using Analysis of Variance (ANOVA). According to the univariate analysis of variance, there was no statistical difference between treatments. There was an interaction effect of teacher and grade. Although writer's workshop was believed to produce writers who have a better self-perception due to differences in time, ownership and response in writing, this study doesn't support the assumption.

For educators this study indicates that regardless of the method of instruction, self-perception does not appear to affect writing achievement. This study however shows that there was no difference in self-perception between these two types of writing instruction. Based on the results of this study, writer's workshop doesn't affect writer's self-perception. Although this study lasted one full school year, questions remain as to whether using these two types of instruction, was long enough to show effects.

Pollington et al... also suggests that longitudinal studies are needed to further investigate the results of this study. In terms of my study, this study is important to consider whereas the results demonstrated here may raise theoretical issues. Specifically speaking, perhaps the writing performance of the participants in my study may be the result of self-perception and not treatment conditions.

In the past two studies, students have rated themselves in terms of writing ability and self-perception. In Stuhlman, J. Daniel, C., Dellinger, A., Denny, K., and Taylor Powers (1999) teachers judge students' writing. The purpose of the study was to investigate teachers' reliability on judging students' portfolio assessment writing using a rubric. This study was a quasi-experimental study. The participants were the teachers. Prior to beginning the target study, a smaller pilot study was conducted to investigate interrater reliability. The treatment group was trained on how to use the rubric to rate portfolio writing, while the control group was untrained. The study took place at two different elementary schools in the same school district. As part of the pilot study, writing portfolios from a first grade classroom in an urban part of the southeastern United States were collected and judged using the writing rubric established over a three-year period by the teachers of that class. Students wrote on the same topic and were judged throughout the year. The same rubric was used to judge student writing in the real study. The rubric used to assess writing consisted of six parts: pre-writing-picture, sentence, punctuation and capitalization, format, story and mechanics. Data was quantitatively measured using an ANOVA in SPSS. The results showed no significant difference between the trained and untrained groups of teachers. In four out of six rubric categories the trained teacher variability was less than half the untrained variability. No differences were detected in the categories of story and mechanics/spelling. All in all, there seemed to be no significant differences in the trained versus the untrained group. Although not highly significant, there was some indication that uniformed training improved reliability. However, internal validity may be called into question whereas random assignment of teachers to groups did not occur.

While Struhlmans' et al (1999) study focused more on holistic scoring, analytic scoring was also widely used. In Roid's (1994) study, assessment issues were discussed comparing holistic scoring to analytic scoring. The purpose of this study was to investigate the advantages of the analytic scoring of writing in grades 3 and 8. This study examined the patterns and validity of analytic scoring using cluster analysis. Random samples were selected for this study. An enormous study of 12,129 third graders and 10,915 eight graders' writing samples from the Oregon Directed Writing Assessment were included. Samples from English as a Second Language (ESOL) and severely disabled students were not included. As for the design of the study, students were randomly assigned to five modes of writing, descriptive, persuasive, expository, narrative, and imaginative. They were allowed three consecutive days, 45 minutes a day, to make and revise a final piece of writing. Within each mode, students had to select from two writing choices. Two trained and one untrained rater scored the drafts using a five point system, low (1) to high (5). Reliability shows 95% agreement. Final writing drafts were scored using these six traits: ideas, organization, voice, word choice, sentence fluency, and conventions. Due to the sample size, 180 students from grade 3 and 100 from grade 8 were involved in the statistical analysis. Using SPSS Quick Cluster Analysis there were 11 clusters types identified. Clusters came from a pilot study conducted in 1985 on eighth grade writing skills. This study only used the descriptive writing mode. This study showed many interesting patterns among the six traits analyzed. Findings suggest that if holistic scores had only been ascertained, patterns within these areas would be missed 60% of the time. The researcher further added that

this six level analytic scoring method did in fact measure the curriculum standards it was designed to measure.

This study raised questions about which scoring method best served the student. Holistic scoring, while popular may not provide as much detail about aspects that would specifically improve writing. The researcher pointed out a need for classroom observations and teacher input on the cluster patterns and whether external writing assessments are consistent with these findings. While holistic scoring can be costly, analytic scoring method was important to consider as a method of measurement.

While the previous few studies focused on the assessment of students' writing, Hooper, S., Swartz, C., Wakely, M., De Kruif, R., and Montgomery, J. (2002) investigated other factors associated with students' writing. The purpose of this quasi-experimental study was to investigate the executive functioning of elementary school students with and without problems with written expression. This study aimed to provide background knowledge on writing as it pertained to a problem solving process. The researchers attempted to compare students without problems in written expression to students with writing problems to determine what factors account for their writing difficulties. The participants in this study included 55 fourth and fifth grade students from two different schools in two different school systems in North Carolina. Numerous and thorough statistics were provided explaining the selection criteria among participants. The independent variables included having participants write two narratives using story starters given by the researcher. After the completion of the narratives, executive function tasks were administered by the same examiner throughout the course of 10 sessions lasting no longer than 30 to 40 minutes. Interrater coefficients were 80%. Each narrative

was scored using National Association of Educational Progress (NAEP) guidelines. Executive functions measurements included the standardized test: Clinical Evaluation of Language Fundamentals (CELF-R), used to measure grammatically correct productive sentence-level verbal expression, Controlled Oral Word Association (COWAT), used to tap verbal organization, retrieval, and general fluency; the Wisconsin Card Sorting Task (WCST), used to measure key executive functions such as problem solving, self-monitoring, and cognitive flexibility; the 4 Tower of Hanoi (TOH 3), used to tap overall problem solving efficiency and self-monitoring; the Word Attack Subtest (WRMT-R), used for reading and decoding problems; and the Child Symptom Inventory (CSI), used to determine the presence of ADHD, MFFT and VSAT (undefined by the researcher). Findings show that there were no significant differences in grade placement, gender, ethnicity, free/reduced lunch, school attended, special education services, or chronological age. However, on the Woodcock Reading Mastery Test, poorer writers scored statistically lower than good writers. Group comparisons revealed that on the executive function domains tested, both groups scored average. Results also suggest that reading decodable words contributed the most to writing outcomes indicating the importance of executive functioning (regulating fluency, putting ideas into text, correcting errors, grammar, spelling and overall monitoring). In terms of narrative writing, word attack significantly contributed to written expression.

This study was not well designed comparing the two groups scores to each other. The researcher hypothesized that the good writers would out perform the poor writers, as would any reader. It was however, unclear what determined a good writer from a poor one. This study involved narrative writing pieces in order to sort participants into good

verses poor writing groups. Neither writing pieces took into account whether participants had opportunities, successful or unsuccessful, with this genre. Multiple writings in different genres would have given more credence to the sorting method. The testing instruments used in the executive function testing were laborious and confusing to read. Understandably, the researchers want to cover their entire basis, but to a reader over use of acronyms and lack of background knowledge caused confusion in the reading. To the researchers and my surprise there were no statistical differences mentioned when comparing “good” verses “poor” writers. This raises the question of whether we can honestly categorize participants this way or is writing ability a continuum? All in all, the topic of this study showed the possibility of identifying the differences in good verses poor writers. The detail and statistics provided were very thorough. In contrast to other studies mentioned, ESOL and students with severe disabilities were excluded. This study is also important because it looks at varying writing abilities.

### Assessing Higher Order Thinking

Hancock’s study (1994) questioned assessment measures. In this quasi experimental study, the purpose was to compare two testing formats to see if there was a difference between multiple-choice questions and constructed response test formats. Hancock chose two intact classes of students in a post baccalaureate program at a university in the northwest. Both classes were given midterm and final exams. Using previous tests and notes from the lectures, the researcher-created several tests according to four levels of Bloom’s taxonomy: knowledge, comprehension, application and analysis. Each test consisted of multiple-choice questions and constructed response

items. Answers to the questions were statistically analyzed. Hancock found that multiple choice and constructed response test formats were not different. Statistically, the results from this study were also calculated according to taxonomic level and, although the difficulty of questions increased, the means did not. Accordingly, this study's findings suggest that the taxonomic of questions cannot be assumed indicative of the difficulty. In terms of the methodology, using two classes added more credibility to the findings. However, as stated, Hancock used previous testing and notes from lecture to create the testing instrument. Hancock's inconsistency in being present to take lecture notes in one class, while not visiting and taking lecture notes in the other class leads to bias in the obtained results. We can only assume that the researchers' choice to use two intact groups reduced testing anxiety within the group but randomization of subjects would have made this study more reliable. Although this was not a true experimental study, it does call into question several things that may be worth investigating. If, as many theorists believe, multiple choice test questions only test the basic recall of information, why were no differences seen in results of the two contrasting question types? Furthermore, as the researcher suggested, what would the findings have shown if Hancock had used standard question stems in the question creation? Additionally, this study casts doubt on the widely accepted belief that Bloom's Taxonomy is a scaffold to increase higher order thinking and cognitive complexity.

### Accountability

Vogler (2002) found that by releasing the Massachusetts Comprehensive Assessment System (MCAS) state mandated test scores to the public, teachers made

significant changes towards teaching using best practice methods. Vogler initially questioned whether the release of state mandated test scores would have an impact on instructional practices of teachers. His 1993 study utilized qualitative and quantitative measures. Through the use of surveys distributed by stratified randomization subparts of the testing instrument were statistically analyzed. This post treatment measure, asked the tenth grade teacher participants to rate whether they increased or decreased the use of twenty instructional strategies, seven teaching techniques, and the use of thirteen instructional materials. The findings suggest that after the public release of test scores, teachers believe they changed their classroom instruction by increasing opportunities for open-ended response questions, relative and critical thinking questions and offering more problem solving strategies. Teachers further indicate that the incidences of lecturing and asking one correct answer type questions have dramatically decreased. Lastly, teachers rate their change in instructional practices directly as a result of teachers' interest in helping students meet the criteria for graduation and helping improve the test results of their school. This is pertinent to the current NCLB high stakes testing, offering hope that as the public becomes more aware of the overall school success rate as determined by NCLB, perhaps teachers will take a more vested interest in improving classroom instruction. Wharton-McDonald's (1998) findings further support this study stating that student achievement often begins with outstanding instruction.

CHAPTER III  
METHODOLOGY  
Research Question

The purpose of this study was to investigate the effects of using Bloom's Taxonomy as an oral questioning scaffold to improve writing in response to reading and reading comprehension through the use of higher order thinking. This study focused on the questions below:

1. Do students demonstrate higher level thinking in their writing in response to teacher questioning from higher levels of Bloom's taxonomy?
2. Do students demonstrate better reading comprehension in response to teacher questioning from higher levels of Bloom's taxonomy?
3. Do students demonstrate higher level thinking in their writing in response to teacher questioning from lower levels of Bloom's taxonomy?
4. Do students demonstrate better reading comprehension in response to teacher questioning from lower levels of Bloom's taxonomy?
5. Does teacher questioning from higher levels of Bloom's taxonomy cause students to demonstrate higher level thinking in writing than does teacher questioning from lower levels of Bloom's taxonomy?
6. Does teacher questioning from higher levels of Bloom's taxonomy improve reading comprehension more than teacher questioning from lower levels of Bloom's taxonomy?

## Setting

This study took place at an average sized elementary school consisting of approximately 500 students located in the suburbs of Washington, D.C. and Baltimore, MD. The study lasted approximately four weeks and occurred during the summer school session from July 1, 2005 to August 1, 2005. The treatment and control conditions lasted for 30 to 40 minutes a day, three out of four days a week. On the fourth day of the week, regular curriculum math, science and social studies instruction took place (instruction not pertaining to this study). The overall study occurred in 12 sessions, totaling 6 – 8 hours of instructional time. This study took place in the regular classroom where participants were comfortable and familiar with surroundings. As the researcher, I visited the classrooms of the participants several times prior to the study so students would recognize and be comfortable with me.

## Pilot Testing

Because I did not have enough participants initially, prior to this study a pilot study was conducted to test the format of using Bloom's Taxonomy as an oral inquiry scaffold, to test the reliability of the writing rubric, the researcher-created test of reading comprehension and DRP test.

The pilot study occurred at a different elementary school in June 2004. 13 students participated but the mortality rate was high, whereas 3 students missed one or two weeks of the four week pilot study. Students were not randomly assigned to groups. Participants ranged from fifth to sixth grade level with the sixth graders having been redistricted and divided between 2 middle schools.

During the pilot study, the format, which was almost identical to the real study, went well. My instruction and materials were the same, including testing instruments. Again, I taught both groups but the classroom was small too small to make sure that the control group was not listening to and or benefiting from the oral inquiry with the experimental group. I learned that I needed a large classroom for my real study. Additionally, attendance was poor. This was evident in all summer school classrooms. Prior to conducting my real study, I asked if the school where I was researching did anything to encourage consistent attendance. During my real research, the school provided all classrooms with weekly snacks to celebrate good attendance.

The results of the pilot caused minor alterations in real study. For instance I taught in a larger classroom for the real study, but other than the above-mentioned changes, the format was the same. Lastly, the pilot study enabled me to be more comfortable when conducting the real research.

### Participants

The participants in this study were 22 fifth grade level students, whose ages ranged from eight to ten years old with the average age 9.5 years old. The teacher to student ratio during this study was 1 teacher per 11 students. The school's ethnic diversity included Caucasian students 62.9%, African American students 22.3%, American Indian or Alaskan Native students 0%, Asian or Pacific Islander students 6.5% and Hispanic students 8.2%. Additionally, 21.1% of the students received free or reduced meals, 6.3% of students had limited English proficiency and 11.4% of students received special education services. I selected this sample of students because 1) all

attended summer academy (summer school program), 2) all scored basic on the 2005 MSA, Maryland School Assessment, indicating below grade level reading proficiency, 3) the administration approved my research and agreed to hire me as the summer school teacher, and 4) the participants were entering grade 5. Because parental consent and student assent were requirements in order to participate in this study, all students were included in the statistical analysis. The participants attended summer academy because the regular classroom teacher recommended it to help the student improve skills, the teacher or parent wanted to decrease the amount of learning lost over the summer, or the student's parent requested they attend. As mentioned, all participants scored basic on the MSA indicating below grade level performance in reading. Other information on the participants' reading level was not available at the onset of the study.

Neither intrinsic nor extrinsic rewards nor reimbursements were offered to the participants before, during or after the study.

### Outcome Variables

This study took place in a summer school setting. Prior to beginning the study, scores from three pre-experimental conditions were collected. Specifically, scores from the Degrees of Reading Power (DRP) test scores from a researcher-created test of reading comprehension, and scores from writing samples were collected. The DRP is a test designed to “inform instruction, monitor a students’ progress toward specific learning goals and provide outcome measures... (TASA Handbook, pg. VII).” It is a standardized test using the cloze method where participants read the selected text and fill in the missing word with one of the four choices provided in the multiple choice format. This

test and the researcher-created test of reading comprehension were used as the pretest and posttest assessment to determine growth in reading.

Additionally prewriting and postwriting samples were collected. The writing samples were letters written by all students in response to the prompt, “using what you know about the text, write a letter to your friend about the story.” The writing samples were compared to an experimenter made rubric (see appendix A). They were scored holistically, and using a point system, to assess higher order thinking as evidenced in writing. The researcher-created rubric in this study is a six level scoring guide where one is the lowest level and six is the highest level. It is comparable to and based on Bloom’s taxonomy. Holistic scoring is the assignment of an overall score based on the leveled rubric. For example, writing that received a holistic score of one, indicates that overall the writing was low level according to Bloom’s taxonomy and involved a lot of duplication of text. Writing that earned a 5 or 6 on the rubric indicates that the participant’s writing overall may have shown evaluation, judgment or justification in their writing comparable to the evaluation level on Bloom’s taxonomy.

Point system scoring is the assignment of points for items shown in the writing. The additional assessor and I scored the writing independently after a brief explanation of how to score the writing. The point scores of writing were assigned first although the mean score, used to statistical analysis, was not calculated until after the holistic score was assigned, to prevent the mean point score of writing possibly influencing the overall holistic score assigned next. To train the additional assessor I modeled how to score the participants writing using a writing sample from the pilot study. When scoring the participant’s writing, the researcher created rubric and writing sample was placed side by

side. The scorer read the participant's work and placed a check mark beside the criteria on the rubric, seen in the participants writing. For instance, to earn 1 point, under the holistic level one, the participant most likely duplicated text or listed information in their writing, as listed under the level one on the rubric. To earn 3 points, the participants' writing content most likely showed a comparison between texts, as listed under the holistic level 3 on the rubric. The total points, which may include 3 1-point checks (3 x 1), 1 2-point checks (1 X 2) and 1 3-point checks (1 x 3) were then averaged to find a mean point score. In this example the total points equaled 8 points divided by 5 checks or criteria seen according to the rubric, for a total of 1.6 averaged points. The average point score was used for analysis and later compared to the holistic score of writing to see if a correlation existed between the holistic and point scored writing. An outside rater also assessed writing ability for interrater reliability although only the researcher's scores were used for analysis. The pretest conditions mentioned (DRP, researcher-created test and writing samples) were used as a baseline to compare changes in reading comprehension and writing as a result of the treatment.

At the end of the study, students took the posttest version of the DRP test form K-9, a researcher-created posttest and a final writing sample.

### Research Design

This research followed the pretest-posttest control group design model. The treatment and control groups were randomly assigned prior to the beginning of summer school 2005. The participants were assigned randomly by listing all participant names on craft sticks, shuffling the craft sticks, then dividing them into two groups; thus reducing

the possibility of non-equivalent of groups. This single blind research occurred in the naturalistic setting of a fifth grade classroom. A standardized test, DRP, a researcher-created test for reading comprehension, and rubrically scored writing samples were the instruments of measurement. Scores were analyzed using t-test and repeated measures Analysis of Variance (ANOVA) for statistical significance.

As the researcher, I was the instructor. I have been an educator for 10 years. I began my teaching experience in 1997 and taught grades kindergarten through third grade for a total of six years. In 2003 I became a middle school reading specialist for two years. In 2005, I was promoted to my current position as middle school assistant principal.

To insure internal validity, I instructed both the control and experimental groups. No participants had previously been students of mine. Because I was the teacher of both groups, one might question whether I provided more assistance to the experimental group than the control group, hence experimenter bias. To ensure fidelity to the treatment, a student teacher intern was present daily and an outside observer visited the classroom once a week to observe my instruction. The outside rater also, scored the writing samples for interrater reliability.

Several conditions were the same for each group yet each group remained intact, in my classroom. While I worked with one group, the student teacher intern monitored the other group in a different part of the classroom. Each day I met with a different group following an AB, BA, AB, and BA pattern to ensure that the same group did not meet with me first daily. Each week both groups worked on the same book that changed weekly totaling 4 trade books covered in 4 weeks. The trade books used weekly were

recommended by Thomas Gunning in *Best Books for Beginning Readers* (1998). The second to third grade level books; *Bread and Jam for Frances* by Russell Hoban, *Ms. Nelson is Missing* by James Marshall, *Stone Soup* by Ann McGovern and *The Paper Crane* by Molly Bang, were chosen to reduce the likelihood that participants would struggle with comprehension when reading thereby decreasing this as a variable allowing the treatment to be investigated. As stated before, the treatment ranged from 30 - 40 minutes a day, three times a week for four weeks, all in all 12 sessions, or 6-8 hours of total instruction. The research followed the schedule below:

#### Instruction that was the Same for Both Groups

Day One: With both groups, I informed students that today they would be listening to and thinking about a story. Later they would be asked to write about the story. After listening carefully, students were given a brief book talk on the book to serve as an introduction. The story was read orally as students followed along in their books.

Day Two: Participants reread the story silently and then with a partner.

Day Three: Participants were asked to reread the text silently. I then gave all participants the following prompt “using what you know about the text, write a letter to your friend about the story.”

This single blind research followed the same format each week, varying only in the reading selection and the oral inquiry method listed below. Each participant had equal opportunity to answer questions asked. While the questioning did not follow a predetermined script, the format of Bloom’s taxonomy was followed for both groups.

### Instruction for Experimental Group

Using question stems based on Bloom's Taxonomy, I asked questions about the weekly story read. With the experimental group, several questions were asked from the top four taxonomic levels (application, analysis, synthesis and evaluation) beginning with the application level and progressing through the hierarchy to the evaluation level.

Day One: After I read the story aloud and students followed along, I asked the experimental group only higher level questions, based on Bloom's taxonomy. I asked the experimental or treatment group questions such as, "justify why...defend... evaluate... interpret..." On occasion, the participants did not understand the meaning of the question stem. After I explained the word, the participants answered the question. Examples included, justify why the restaurant owner treated the stranger to a meal. Defend the restaurant owner's decision to keep the restaurant open after the highway was built. Explain the significance of the paper crane.

Day Two: After rereading the story twice, I asked the experimental group only higher level questions, based on Bloom's taxonomy.

Day Three: After a fourth reading, I asked the experimental group only higher level questions, based on Bloom's taxonomy.

### Instruction for Control Group

Day One: After I read the story aloud and students followed along, I asked the control group lower level questions from Bloom's knowledge and comprehension levels. Examples included, "Who were the characters in the Paper Crane? How many characters

were in the story? Where did the Paper Crane take place? When did the story take place? List what happened in the story to the boy in the story.”

Day Two: After rereading the story twice, I asked the control group lower level questions, based on Bloom’s taxonomy.

Day Three: After a fourth reading, I asked the control group lower level questions based on Bloom’s taxonomy. Participants were asked to picture the story in their head and draw a picture about the story.

I taught both groups and did not want the experimental group to know that they were the group receiving the treatment. To counteract the Hawthorne Effect, participants in the control group were asked to picture the story in their head and draw a picture about the story. Participants from the experimental group asked why they could not draw a picture too, and wondered why the control group got to do a special activity. This response reassured me that neither group knew which group was receiving the treatment.

### Analysis of Data

Performances scored were the pretest and posttest DRP, researcher-created test and the pre-test/posttest writing samples. T-test and ANOVA were the statistical tests administered to compare the control group to the treatment group. The probability level for the test was  $<0.05$ . The hypothesis, students’ writing in response to reading and reading comprehension will improve as a result of oral inquiry scaffolding of texts, based Bloom’s taxonomy. Only the researcher’s scores were used in analysis.

## Projected Qualitative Analysis of Data

Qualitatively speaking, notes were taken throughout the study to informally assess the verbal use of higher order thinking. For example, when the control group was asked “who were the characters in the Paper Crane?” A participant responded, “the restaurant owner, stranger, boy and dancing crane.” When asked, “list what happened in the story.” A participant responded, “the restaurant was busy, then a new road was built far from the restaurant, no one came anymore, a stranger came in who was hungry but had no money, the owner fed him, he left a paper crane, it danced, more people came to the restaurant.”

When the experimental group was asked, “Justify why the owner treated the stranger to a meal.” A participant responded, “The Japanese are kind people. The owner wanted to be kind even though he was getting poor with no customers. He fed the stranger to be nice probably hoping that the stranger would go and tell people how nice the owner and restaurant was and more people would come and he wouldn’t be poor anymore.” Another participant responded, “Maybe he thought the stranger would love the food and go tell others to come. Then the owner would have more money.” Lastly, a participant responded, “Maybe the owner believes that you should treat others like you want them to treat you.”

CHAPTER IV  
DATA ANALYSIS AND RESULTS

Study Overview

The purpose of this study was to investigate the effects of Bloom's Taxonomy as an oral questioning scaffold to improve writing in response to reading and reading comprehension through the use of higher order thinking.

This study focused on the questions below:

1. Do students demonstrate higher level thinking in their writing in response to teacher questioning from higher levels of Bloom's taxonomy?
2. Do students demonstrate better reading comprehension in response to teacher questioning from higher levels of Bloom's taxonomy?
3. Do students demonstrate higher level thinking in their writing in response to teacher questioning from lower levels of Bloom's taxonomy?
4. Do students demonstrate better reading comprehension in response to teacher questioning from lower levels of Bloom's taxonomy?
5. Does teacher questioning from higher levels of Bloom's taxonomy cause students to demonstrate higher level thinking in writing than does teacher questioning from lower levels of Bloom's taxonomy?
6. Does teacher questioning from higher levels of Bloom's taxonomy improve reading comprehension more than teacher questioning from lower levels of Bloom's taxonomy?

A true experimental pretest and posttest control group design was used. Both the control and treatment groups were randomized to make this a true experimental study. This study attempted to determine if teacher questioning, using higher order question stems based on Bloom's taxonomy would improve students' writing in response to reading and reading comprehension. Student outcomes were measured in four ways: the DRP test (forms J-9 and K-9), a researcher-created reading comprehension test using higher order questions and holistic and point scoring of writing using a researcher-created rubric influenced by Bloom's Taxonomy. Paired T-Test, Pearson correlation coefficients and a 2 X 2 (repeated measures) ANOVA statistical analysis were used to analyze data.

Before investigating the statistical data used to answer each question, it is important to note that an independent samples T-test compared the control and experimental group's pretest scores on holistic writing, point scored writing, DRP test of reading comprehension and the researcher created test of reading comprehension. The results of the independent samples T-test showed that there were no significant differences in pretest scores from the two groups.

#### Findings Related to Question #1

The goal of research question # 1 was to investigate whether students demonstrate higher-level thinking in their written response to teacher questioning from higher levels of Bloom's taxonomy?

To assess students' writing, scores were calculated two ways, holistically and using a point system. Holistic scoring assessed the participants' overall written work

using a leveled rubric. The researcher-created, six level rubric was designed based on Bloom’s taxonomy which was also used as the oral inquiry scaffold. While the rubric is divided into six levels for holistic scoring, the sixth being the highest, each level also consisted of multiple items supporting that level. In contrast to the holistic score, the participants’ content writing earned points for displaying items listed in each rubric level. Then, point averages were used to score writing. Later in this chapter, the relationship between holistic scoring and point scoring will be investigated.

Holistic and point scores were assessed week one (as a pretest), prior to use of the treatment and then were compared to week four (posttest), at the end of the study, to determine the effects of the treatment on the control and experimental groups’ writing skills.

**Table 1 Paired T-Test for the Experimental Group Holistic Scores Pretest and Posttest**

---

	N	M	SD	t	p
Experimental Group Holistic Scores					
Pretest	11	2.23	0.90		
Posttest	11	4.55	0.91		
Difference	11	2.32	1.01	7.64	< 0.05

95% CI for mean difference: (1.64, 2.99)  
T-Test of mean difference = 0 (vs not = 0)

---

A paired T-Test was used to analyze the change in the experimental group’s mean holistic writing score on the pretest and posttest. A paired T-Test was used to determine whether significant differences in the pretest and posttest mean scores were seen in each group. Of the 11 participants assessed, the results showed that there was a significant increase in the holistic writing scores of the experimental group from week one to four.

The scores on this measurement can range from one to six with a score of one indicating a low holistic level, comparable to the knowledge level on Bloom’s taxonomy, and a six score, indicating a high holistic level, comparable to the evaluation level on Bloom’s taxonomy. Table 1 shows a  $p < .05$ , which is statistically significant at the 0.05 level, indicating that the experimental group significantly improved their holistic writing scores. Considering the mean difference scores, this indicates that the mean improvement for any student would be between 1.64 and 2.99.

Table 2 Paired T-Test for the Experimental Group Point Scores Pretest and Posttest

---

	N	M	SD	t	p
Experimental Group Point Scores					
Pretest	11	2.48	1.04		
Posttest	11	4.96	0.97		
Difference	11	2.49	0.97	8.53	< 0.05

95% CI for mean difference: (1.84, 3.14)  
T-Test of mean difference = 0 (vs not = 0):

---

The second measurement of writing used the point scoring method previously discussed. While the average point score ranged from one to six, one is the lowest average point earned, comparable to the knowledge level on Bloom’s taxonomy. Six is the highest average point earned, comparable to evaluation of Bloom’s taxonomy. According to Table 2, there was a significant difference between the point scores of writing from weeks one (pretest) to four (posttest) with the  $p < .05$ , which is statistically significant at the 0.05 level. The mean differences in score were positive indicating that, according to this study, the students improved their writing with the mean improvement ranging from 1.84 to 3.14.

Pearson correlation coefficient showed that the holistic and point scoring methods were equivalent for the experimental group during weeks one and four. Correlations coefficient ranged from 0.75 to 0.88 and a  $p < .01$ , statistically significant at the 0.05 level.

### Findings Related to Question #2

The goal of research question # 2 was to investigate whether students demonstrate better reading comprehension in response to teacher questioning from higher levels of Bloom’s taxonomy? Reading comprehension was measured in two ways, DRP standardized test and a researcher-created test of reading comprehension using higher order question stems. The DRP scores can range from one to 42; one, the lowest score and 42, the highest score

Table 3 Paired T-Test for the Experimental Group DRP Test of Reading Comprehension Pretest and Posttest

---

	<u>N</u>	<u>M</u>	<u>SD</u>	<u>t</u>	<u>p</u>
Experimental Group DRP Scores					
Pretest	11	25.73	7.59		
Posttest	11	30.64	8.13		
Difference	11	4.91	7.89	2.06	0.06

95% CI for mean difference: (-0.39, 10.21)  
T-Test of mean difference = 0 (vs not = 0)

---

The results of the DRP test comparing the pretest form J-9 to the posttest form K-9, were analyzed using a paired t-test. The data (Table 3) showed that there was no significant difference in scores with the  $p = 0.06$ . While the p value is close to the  $p <$

0.05 level, oral inquiry using Bloom’s taxonomy may improve the participants’ reading comprehension although the improvement is not significant in scientific terms. The mean difference DRP comprehension scores were negative and positive, indicating that the mean improvement for the participants in the experimental group fell between –0.39 and 10.21. All in all this means that according to this study, sometimes, reading comprehension was unchanged and sometimes growth in DRP reading comprehension scores were seen.

Table 4 Paired T-Test for the Experimental Group Researcher Test of Reading Comprehension Pretest and Posttest

---

	N	M	SD	t	p
Experimental Group Researcher Test Scores					
Pretest	11	5.36	2.01		
Posttest	11	7.27	2.10		
Difference	11	1.91	2.34	2.70	< 0.05

95% CI for mean difference: (0.33, 3.48)  
T-Test of mean difference = 0 (vs not = 0)

---

The other test of reading comprehension was the researcher-created test. This 13-question test was created using higher order questioning stems from Bloom’s taxonomy. Scoring on the test ranged from one to 13 with one, the lowest score and 13, the highest. While 11 participants were assessed, the results in Table 4 indicates that the experimental group made significant improvements in reading comprehension on pretest and posttest mean scores on the researcher created test of reading comprehension The data below reports a  $p < 0.05$ , which is statistically significant, at the 0.05 level. In addition, both

mean difference scores were positive with the mean improvement ranging from 0.33 to 3.48.

### Findings Related to Question #3

The goal of research question #3 was to investigate whether students demonstrate higher-level thinking in their written response to teacher questioning from lower levels of Bloom’s taxonomy?

Table 5 Paired T-Test for the Control Group Holistic Scores Pretest and Posttest

---

	<u>N</u>	<u>M</u>	<u>SD</u>	<u>t</u>	<u>p</u>
Control Group Holistic Scores					
Pretest	10	2.00	1.18		
Posttest	10	1.95	0.60		
Difference	10	-0.05	1.12	-0.14	0.89

95% CI for mean difference: (-0.85, 0.75)  
T-Test of mean difference = 0 (vs not = 0)

---

Considering the range in holistic scores of writing was one (the lowest) to six (the highest), Table 5 shows that the control group did not make significant gains in writing when scored holistically. The p value = 0.89, not significant at the 0.05 level. Table 5 shows the 95 % confidence interval mean differences were between –0.85 and 0.75. Negative and positive numbers indicate that sometimes the scores improved, sometimes holistic writing scores did not change. All in all, no significant differences were seen in the holistic writing scores from the control group.

Table 6 Paired T-Test for the Control Group Point Scores Pretest and Posttest

---

	N	M	SD	t	p
Control Group Point Scores					
Pretest	10	2.35	1.14		
Posttest	10	2.17	0.78		
Difference	10	-0.18	0.73	-0.79	0.45

95% CI for mean difference: (-0.70, 0.34)

T-Test of mean difference = 0 (vs not = 0)

---

The second measurement assessed writing using the point system. According to table 6, no significant difference in the mean point score of writing was observed. The p value = 0.45, which was not statistically significant, at the 0.05 level. The 95% confidence interval for mean difference shows a negative and positive number indicating that sometimes the mean point score writing improved, and sometimes no change in writing was seen. The mean improvement ranges from -.70 to 0.34. All in all no significant differences in point scored writing was seen during this study.

Pearson correlation coefficient compared the holistic and point scoring method where data indicated that the two methods were very similar with correlations ranging from 0.70 to 0.74 for weeks one and four. P values < 0.01, which is statistically significant, at the 0.05 level. According to these statistics, there was a relationship between scoring the control group's writing holistically and with the point system.

#### Finding Related to Question #4

The purpose of question #4 was to investigate whether students demonstrate better reading comprehension in response to teacher questioning from lower levels of

Bloom's taxonomy? To investigate this, the participants took two tests, DRP and a researcher-created test, using higher order questioning stems, to measure reading comprehension.

Table 7 Paired T-Test for the Control Group DRP Test of Reading Comprehension Pretest and Posttest

---

	<u>N</u>	<u>M</u>	<u>SD</u>	<u>t</u>	<u>p</u>
Control Group DRP Scores					
Pretest	11	24.73	6.51		
Posttest	11	24.82	9.63		
Difference	11	0.09	6.46	0.05	0.96

95% CI for mean difference: (-4.25, 4.43)  
T-Test of mean difference = 0 (vs not = 0)

---

Table 7 shows the results from the control group, on the DRP test. The p value = 0.96, which is not statistically significant at the 0.05 level, shows that there is no significant change in the control group's mean reading comprehension score on the DRP test. Considering that the scores on the DRP can range from one to 42, the 95% confidence interval mean difference shows that the mean improvement ranged from -4.25 to 4.43. Therefore according to this study, when participants were taught using low-level oral inquiry, sometimes their reading comprehension improved and sometimes there was no change. All in all, the control group did not make significant gains based on the DRP test results.

Table 8 Paired T-Test for the Control Group Researcher Test of Reading Comprehension Weeks One and Four

---

	<u>N</u>	<u>M</u>	<u>SD</u>	<u>t</u>	<u>p</u>
Control Group Researcher Test Scores					
Pretest	11	6.55	2.88		
Posttest	11	7.09	2.91		
Difference	11	0.55	2.38	0.76	0.47

95% CI for mean difference: (-1.05, 2.15)  
T-Test of mean difference = 0 (vs not = 0)

---

The second assessment instrument used to measure reading comprehension was the researcher-created test using higher order questions stems based on Bloom’s taxonomy. According to Table 8, there was no significant difference in the reading comprehension mean scores on the researcher-created test of reading comprehension. The p value = 0.47, which was not statistically significant, at the 0.05 level. This two tailed test, shows a mean difference indicating that when participants were taught using lower level oral inquiry, sometimes their reading comprehension scores improved and sometimes there was no change in reading comprehension scores according to the researcher-created test. The mean improvement ranged from –1.05 to 2.15. While the researcher-created test was based on higher level questioning stems from Bloom’s taxonomy, the control group did not make significant gains in reading comprehension scores.

### Findings Related to Question #5

The goal of question # 5 was to investigate whether teacher questioning from higher levels of Bloom’s taxonomy caused students to demonstrate higher level thinking in writing than did teacher questioning from lower levels of Bloom’s taxonomy?

Levene’s test of equality of error variances showed that there were differences in variance of the holistic scores. For the pretest holistic score the  $p = 0.78$ , and the posttest holistic the  $p = 0.44$ , neither statistically significant, at the 0.05 level. It was important to investigate whether differences existed before investigating whether the differences were statistically significant.

Table 9 Descriptive Statistics for the Between Subjects Pretest Posttest Holistic Scores of the Control and Experimental Group

	<u>Group</u>	<u>N</u>	<u>M</u>	<u>SD</u>
Pretest Holistic Scores	Control	10	2.00	1.18
	Experimental	11	2.23	0.90
	Total	21	2.12	1.02
Posttest Holistic Scores	Control	10	1.95	0.60
	Experimental	11	4.55	0.91
	Total	21	3.31	1.53

Table 9 shows descriptive statistics for the groups. With a possible range of one, being the lowest, to six, the highest score; the mean score from both groups were comparable with a 2.00 for the control group pretest and 2.23 for the experimental group’s pretest. Posttest scores showed a difference between the control group’s mean score 1.95, and the experimental group’s posttest mean score, 4.55.

Table 10 Repeated Measure Analysis of Variance for the Pretest Posttest Holistic Scores of the Control and Experimental Groups

Source	Type 3 Sum of Squares	df	F	p	Partial eta Squared
Group	20.87	1	18.42	< 0.05	0.49
Holistic	13.47	1	23.97	< 0.05	0.56
Holistic*Group	14.69	1	26.13	< 0.05	0.58

Table 10 shows the between and within subject effects of group assignment and holistic writing score. A significant interaction emerged between the holistic writing scores and the random group assignment; the p value < 0.05, statistically significant at the 0.05 level. The medium effect size of 0.58, (partial eta squared) of the interaction between the holistic score indicates that the group that students were randomly assigned to was a significant predictor of their posttest writing score. Four multivariate tests were performed between the groups corroborating these results.

When considering the point scoring of writing, Levene's test of equality of error variances showed that differences existed in pretest and posttest point scores. The pretest p = 0.51 and the posttest p = 0.63, neither of which was statistically significant at the 0.05 level.

Table 11 Descriptive Statistics for the Between Subjects Pretest Posttest Point Scores of the Control and Experimental Group

	Group	N	M	SD
Pretest Point Scores	Control	10	2.41	1.17
	Experimental	11	2.48	1.04
	Total	21	2.44	1.08
Posttest Point Scores	Control	10	2.17	0.78
	Experimental	11	4.96	0.97
	Total	21	3.63	1.67

Table 11 shows the data from the pretest and posttest point writing scores from the control and experimental groups. The pretest means were comparable, 2.41 for the control group and 2.48 for the point scores of the experimental group. Posttest scores were different with 2.17 for the control group and 4.96 point scores for the experimental group.

Table 12 Repeated Measure Analysis of Variance for the Pretest Posttest Point Scores of the Control and Experimental Groups

Source	Type 3 Sum of Squares	df	F	p	Partial eta Squared
Group	21.54	1	13.33	< 0.05	0.94
Point	31.25	1	34.38	< 0.05	0.64
Point*Group	19.43	1	19.43	< 0.05	0.73

Table 12 shows the between and within subjects effects for the point scores of the control and experimental groups. The data shows a significant difference exists between the point scores of the control and experimental groups with a  $p < 0.05$ , statistically

significant at the 0.05 level. The effect size, partial eta squared, is 0.73 which is medium to large effect size suggesting that the posttest point writing scores are strongly predicted by the random assignment of the participants to groups. Four additional multivariate tests were performed with the same results indicated.

All in all question # 5 investigated whether teacher questioning from higher levels of Bloom's taxonomy caused students to demonstrate higher level thinking in writing than did teacher questioning from lower levels of Bloom's taxonomy? The repeated measures ANOVA, showed that instruction using a higher order questioning scaffold does cause higher level writing through holistic and point scoring, than when participants were instructed using lower level questioning.

Critics have argued that non-standardized testing, including holistic and point scoring, is subjective and varies with the assessor. To investigate this concern, an additional assessor also read the participants' writing and scored it holistically and using the points system. Pearson correlation coefficients compared the additional assessor's scores against the researcher's scores to determine interrater reliability. Interrater reliability was established to investigate whether the scoring of writing was subjective. Pearson correlation coefficient indicates that there was a significant correlation between all interrater measures with all p values < 0.05. Correlation coefficients were strong with the researcher and additional assessor's holistic scores ranging from 0.96 to 0.98. The point scored writing assessed by both scorers had correlation coefficients ranging from 0.93 to 0.99. Whether the outcomes pertained to holistic scoring or point scoring, with the control group or experimental group, statistical significance was achieved. Regardless of the assessor, scores were equivalent on the participants writing. All

interrater reliability correlations were significant and high indicating that, according to this study, the holistic and point scoring of writing was not subjective. Therefore according to this study, both assessors, had comparable scores every week, whether scoring participants' writing holistically or using a point system, with little variability amongst the two scorers. It appears that subjectivity of scoring, when pertaining to student writing, was not a factor affecting the overall conclusion.

#### Findings Related to Question #6

The goal of question #6 was investigate whether teacher questioning from higher levels of Bloom's taxonomy improve reading comprehension more than teacher questioning from lower levels of Bloom's taxonomy? Reading comprehension was measured with the Degrees of Reading Power (DRP) test and a researcher-created test of reading comprehension. The DRP test was investigated first.

Levene's test of equality of error variances showed that there were differences in variance of the reading comprehension scores on the Degrees of Reading Power, DRP test. For the pretest DRP score, the  $p = 0.22$ , and the posttest DRP score, the  $p = 0.27$ , with neither being statistically significant, at the 0.05 level. It was important to investigate whether differences existed before investigating whether the differences were statistically significant.

Table 13 Descriptive Statistics for the Between Subjects Pretest Posttest Scores on the DRP test for the Control and Experimental Groups

	Group	N	M	SD
Pretest DRP Scores	Control	11	24.73	6.51
	Experimental	11	25.73	7.59
	Total	22	25.23	6.92
Posttest DRP Scores	Control	11	24.82	9.63
	Experimental	11	30.64	8.13
	Total	22	27.73	9.19

Table 13 shows the descriptive statistics for the pretest-posttest scores on the DRP test for the control and experimental groups. The DRP scores can range from one, the lowest score to 42, the highest score. On the pretest, there was little variance in the scores with the control group mean 24.73 and the experimental group mean 25.73. The posttest scores showed larger differences with the control group mean 24.82 and the experimental group mean 30.64.

Table 14 Repeated Measure Analysis of Variance for the Pretest Posttest DRP Scores of the Control and Experimental Group

Source	Type 3 Sum of Squares	df	F	p	Partial eta Squared
Group	127.84	1	1.24	0.06	0.58
DRP	68.75	1	2.65	0.12	0.12
DRP*Group	63.84	1	2.46	0.13	0.11

Table 14 shows the between and within subjects effects for the DRP of the control and experimental groups. The data shows that a significant difference does not exist between the DRP scores of the control and experimental group with a  $p = 0.13$ , which is

not statistically significant at the 0.05 level. The effect size, partial eta squared, is a weak 0.11 indicating that the magnitude of the random group assignment affecting the DRP score is minimal. Four additional multivariate tests were performed with the same results indicated.

The second measurement of reading comprehension was the researcher-created test. On the 13-question test, Levene’s test of equality of error of variances showed that differences exist in the scores of the control group and experimental group on the researcher-created test. Results show a pretest score  $p = 0.25$  and posttest score  $p = 0.19$ , with neither being significant at the 0.05 level.

Table 15 Descriptive Statistics for the Between Subjects Pretest Posttest Scores on the Researcher-created test for the Control and Experimental Group

	Group	N	M	SD
Pretest Researcher Scores	Control	11	6.55	2.88
	Experimental	11	5.36	2.01
	Total	22	5.95	2.50
Posttest Researcher Scores	Control	11	7.09	2.91
	Experimental	11	7.27	2.10
	Total	22	7.18	2.48

Table 15 shows descriptive statistics for the control and experimental group on the researcher-created test of reading comprehension. The scores on this test can range from one, the lowest score, to 13, the highest score. According to table 15, the mean on the pretest for the control group was higher ( $M = 6.55$ ) than the mean from the experimental group ( $M = 5.36$ ). On the posttest, the control group mean was 7.09 while the experimental group mean was 7.27.

Table 16 Repeated Measure Analysis of Variance for the Pretest Posttest Researcher-created Test Scores of the Control and Experimental Groups

Source	Type 3 Sum of Squares	df	F	p	Partial eta Squared
Group	2.75	1	0.28	0.60	0.01
Researcher-created	16.57	1	5.94	0.02	0.23
Researcher-created *Group	5.11	1	1.82	0.19	0.08

Table 16 shows the between and within subjects effects on the researcher-created test of reading comprehension on the control and experimental groups. The data shows that a significant difference does not exist on the researcher-created test scores of the control and experimental groups with a  $p = 0.19$ , which is not statistically significant at its 0.05 level. The effect size, partial eta squared, was very small at 0.08 indicating that the random group assignment does not have a significant effect on the reading comprehension scores on the researcher-created test. Additional multivariate tests corroborated the same results.

### Summary

All in all, when investigating participant's writing in response to reading and reading comprehension, two tailed, paired T-test, repeated measures ANOVA and Pearson correlation coefficient tools were used to analyze the data from this study. The experimental group's mean scores on holistic and point scored writing improved significantly from week one (pretest) to week four (posttest). The control group's data did not indicate significant changes in mean score of holistic or point scored writing.

Strong effect sizes further suggest that there was a significant interaction between the random group assignment of participants and the writing scores earned. The overall mean scores in reading comprehension were inconclusive from the control and experimental groups. Weak effect sizes failed to show an interaction between the control group scores on the DRP and researcher-created test. Overall, significant evidence suggests that significant gains in writing were seen as a result of the teacher instructing participants using a higher order oral questioning scaffold based on Bloom's taxonomy.

## CHAPTER V

### DISCUSSION

#### Conclusion

The driving force behind No Child Left Behind (NCLB), currently acting as the undercurrent of education, makes the results of this study important. While NCLB places great weight on assessment results, teachers would benefit from strategies that would help students be successful today and longitudinally. This research involved many aspects that affect students in all classrooms today. Higher order questioning, oral inquiry, scaffolding, writing assessment, and accountability all play important parts in this research.

The goal of this study was to investigate the effects of using Bloom's taxonomy as an oral questioning scaffold to improve writing in response to reading and reading comprehension by encouraging higher order thinking. This study followed a pretest-posttest control group design. Participants, 22 fifth-grade students from a suburban school, were randomly assigned to control and experimental groups. The data was analyzed using a two tailed paired T-test, repeated measures ANOVA and Pearson correlation coefficient.

#### Interpreting Data

The purpose of this study was to investigate the effects of using Bloom's Taxonomy as an oral questioning scaffold to improve writing in response to reading and

reading comprehension through the use of higher order thinking. The research questions investigated were:

1. Do students demonstrate higher level thinking in their writing in response to teacher questioning from higher levels of Bloom's taxonomy?
2. Do students demonstrate better reading comprehension in response to teacher questioning from higher levels of Bloom's taxonomy?
3. Do students demonstrate higher level thinking in their writing in response to teacher questioning from lower levels of Bloom's taxonomy?
4. Do students demonstrate better reading comprehension in response to teacher questioning from lower levels of Bloom's taxonomy?
5. Does teacher questioning from higher levels of Bloom's taxonomy cause students to demonstrate higher level thinking in writing than does teacher questioning from lower levels of Bloom's taxonomy?
6. Does teacher questioning from higher levels of Bloom's taxonomy improve reading comprehension more than teacher questioning from lower levels of Bloom's taxonomy?

The questions aren't as simple as they seem. In this study, students' writing was holistically scored against a researcher-created six level rubric and through a point system where students earned points for addressing the criteria listed in the six level rubric. For instance, if a student lists information verbatim from the text in their writing, they earn one point because verbatim repetition of text is listed under level one on the rubric (see Appendix A). If a student makes an evaluation or judgment about the text then they earn

six points because this criterion is listed under level six on the rubric. The overall point score is found by determining the point score mean. It was also important to investigate if there was a relationship between holistic and point scoring.

Question # 1 asks, do students demonstrate higher level thinking in their writing in response to teacher questioning from higher levels of Bloom's taxonomy? Pretest and posttest scores were compared using a paired t-test, two tailed. The experimental group received the treatment instruction where instruction followed a higher level oral questioning scaffold based on Bloom's taxonomy. The question words used with the experimental group were at application, analysis, synthesis and evaluation levels of Bloom's taxonomy. The results show that students demonstrate higher-level thinking in their written response to higher-level oral inquiry from the teacher. The p value  $< 0.05$ , at the 0.05 level. The 95% confidence interval mean difference shows that this treatment significantly improved the mean score on students' writing an average of 1.64 to 2.99 out of six rubric levels. This is a large gain indicating that 6-8 hours of instruction can increase writing scores significantly. Educators should consider this strong gain and investigate the effects of longer instruction using this treatment. In this study assessment through writing allowed me to evaluate the participants' thoughts and reasoning in writing, not merely using traditional testing to see who has the right answer (Graves, 2000). Dewey (1902) would further agree that what the participants do with what they have read is the essence of thinking and learning.

To complement the holistic scoring data, there were significant improvements noted in the point scores of the experimental group. The paired t-test results indicate a p value  $< 0.05$ , at the 0.05 level. The mean difference indicates that when this treatment

was used, mean writing scores improved with average improvements 1.84 to 3.14 points on a six-point scale. These results further suggest that educators may help students improve writing content when they're taught using this scaffold. Additionally, significant correlations of 0.75 to 0.88 between holistic and point scoring indicate the two scoring measures are interchangeable.

This seems to contradict Hancock's (1994) doubt that higher order questioning improves higher-level thinking. While Hancock did not find differences in multiple choice and higher level constructed response type questioning, this study found that when the participants were able to construct their own responses about text, higher level content was present. Not only does the data from question #1 show differently, but I suggest that the participants had to develop more complex thought and higher level thinking to be able to express higher level thinking in their writing. When children can recreate and reinvent what they encounter in an open format when tested, then assessment is considered authentic (Zemelman, et al., 1998). In this study, the participants were able to use their own words, language, and voice, to express their thoughts individually. Zemelman, et al., (1998) would agree that writing in this study was authentic assessment. Further convincing were the noted gains considering there were only 11 participants in each group. The holistic and point-scoring mean improvement range indicated that some participants doubled their score from weeks one to four. This was seen in the mean improvement 2.99 (holistic top improvement) and 3.14 (point top score improvement). The participants who gained roughly three points improved their score greater than 100% in four weeks. This rate of improvement is impressive and suggests higher order questioning is beneficial.

Question #2 asks whether students demonstrate better reading comprehension in response to teacher questioning from higher levels of Bloom's taxonomy. Reading comprehension was measured using Degrees of Reading Power (DRP) test and a researcher-created test of reading comprehension. These two instruments differ. The DRP test is a standardized test requiring students to choose one of four provided words to complete a sentence, the cloze method. This multiple-choice format does not use higher order questioning stems like the researcher-created test of reading comprehension. In contrast to question #1, where the top score when assessing writing was the same (a score of 6), the DRP top score is 42 and the researcher-created test top score is 13.

A paired t-test shows that when students are taught using higher level questioning based on Bloom's taxonomy, no significant difference in mean score is shown in the DRP scores. The p value of 0.06 was not statistically significant at the 0.05 level. While the p value is very close to the 0.05 level, some improvements are noted, although not significant. The DRP scores can range from 1 to 42. While the mean difference is 4.91, compared to the number of questions (42) this improvement is not significant. The mean difference shows that sometimes reading comprehension improved and sometimes there was no change in reading comprehension. Though statistical significance is not seen, the mean improvements range from -0.39 to 10.21. Knowing that this is the average improvement, it appears that stronger gains in reading comprehension were seen than scores that regressed. Perhaps the lack of significant was due to the participants' loss of interest in the test, they're not putting forth their best effort, growing fatigued, not feeling well or not caring how well they did when taking this 42-question test. This was the longest of the scoring instruments. Alternatively, the relatively brief treatment was

probably not powerful enough to create general improvement in reading comprehension on a standardized test.

On the 13-question researcher-created test, significant gains in mean score were seen when participants were taught using the higher level-questioning scaffold. The paired t-test results show the p value  $< 0.05$ , which is statistically significant at the 0.05 level. Mean improvements ranged from 0.33 to 3.48. In contrast to the 42-question DRP, this 13 question created test probably did not cause testing fatigue. Additionally, the experimental groups had 6-8 hours of instruction using higher level questions. Perhaps when participants took the researcher-created test, the language of the questions stems were similar to the question stems used on the test. All question stems during instruction and on the researcher-created test came from Bloom's taxonomy.

These results are encouraging for several reasons. Some researchers claim that standardized assessments, traditionally with multiple choice type questions, are the only valid measurements of skills. Other researchers believe that most multiple choice formatted questions ask only low level questions, and that authentic assessment requires opened-ended questions. Critics of multiple-choice tests may need to reconsider their rejection of multiple-choice tests and consider that this assessment format can feature higher-level questions if carefully constructed. All in all, do students demonstrate better reading comprehension in response to teacher questioning from higher levels of Bloom's taxonomy? The answer is inconclusive but encouraging. Contradictory results were reported on the DRP and researcher-created test of reading comprehension. While the researcher-created test did indicate significant mean score gains on the paired t-test, the DRP gains were not significant. When considering the overall question the results from

this study find inconclusive evidence that reading comprehension improves when participants are questioned using higher order question stems based on Bloom's taxonomy

Question #3 asks whether students demonstrate higher level thinking in their writing in response to teacher questioning from lower levels of Bloom's taxonomy. According to the results, lower level questioning does not improve the mean score of writing from the participants studied. The p value was 0.89, which is not statistically significant at the 0.05 level. When participants are questioned using low level questions, sometimes the writing score improved; sometimes there was no change in the participants writing score. The improvement range was  $-0.85$  to  $0.75$ .

Using low-level questions did not improve the mean score of writing in response to reading. The p value was 0.45, not statistically significant at the 0.05 level. Again the 95% confidence interval mean difference was negative and positive, indicating that sometimes low level questioning improves writing content. Sometimes there was no change in the mean score of writing. Improvements ranged from  $-0.70$  to  $0.34$ .

The relationship between the holistic and point scoring of the control group's writing was significant. The Pearson correlation coefficients were statistically significant at the 0.05 level. Correlations were 0.74 for week one and 0.70 for week four scoring. This indicates while the scoring of writing holistically and using the point system are not identical, they are closely related.

After instruction the control group was asked to picture the story in their head and draw a picture about it. This was an intentional strategy to counteract the Hawthorne Effect. Based on the holistic scoring results, consequently, using imagery and drawing

did not significantly improve the writing content mean scores either. Possibly when students are asked lower level questions, attention was drawn to the memorization of facts instead of deeper thought about the meaning. When students are then asked to write about the story, the simple recounting of facts on paper is the easiest to produce.

Do students demonstrate higher-level thinking in their writing in response to teacher questioning from lower levels of Bloom's taxonomy? The answer is no. According to the results of this study, lower level oral inquiry does not statistically improve higher level thinking, as measured by the mean scores of participants' writing.

Question # 4 asks whether students demonstrate better reading comprehension in response to teacher questioning from lower levels of Bloom's taxonomy. The results from the DRP test and a researcher-created test showed that low level oral inquiry does not significantly improve the reading comprehension mean score on the DRP test. The paired T-test shows a p value = 0.96, which is not statistically significant at the 0.05 level. On the 42-question DRP test, improvement ranged from -4.25 to 4.43, indicating that sometimes lower level questioning improves the mean score of reading comprehension. Sometimes there was no change in reading comprehension. Perhaps the participants may have lost interest in this test, didn't put forth their best effort, were fatigued, were not feeling well, or didn't care how well they did on this 42-question test. Perhaps the participants had never taken a test with cloze sentences. This information suggests lower level questioning does not help students achieve greater reading comprehension on standardized tests.

The 13 question, researcher-created multiple-choice test used higher order question stems based on Bloom's taxonomy. The results showed that lower level

questioning did not significantly improve the mean score of reading comprehension according to the results of this study ( $p = .47$ ). Improvements ranged from negative (-1.05) to positive (2.15), indicating that sometimes the mean score of reading comprehension improved and sometimes there was no change in reading comprehension. Students were instructed using low-level oral inquiry requiring students to recall facts and restate information verbatim. The researcher-created test asked question stems from Bloom's synthesis and evaluation levels. Perhaps when the participants took this assessment, they were not used to the thought processes and language of the questioning stems necessary to answer correctly. Results suggest we cannot expect to see higher level thinking, orally or in writing, when students are not taught using higher level questioning as a scaffold.

Question #5 asks whether teacher questioning from higher levels of Bloom's taxonomy caused students to demonstrate higher level thinking in writing than did teacher questioning from lower levels of Bloom's taxonomy. This question is critical for determining the effectiveness of the innovative questioning used in this experiment.

Levene's test of equality of variances showed that the control and experimental groups' holistic and point scores differed. When considering the control and experimental group's total score, unrelated (between subjects), the results clearly indicate significant improvement in writing when students are taught using higher level oral inquiry. When the scores were matched for within subject comparison, the results clearly showed that there was an interaction effect between the writing scores and the group assignment. Repeated measure ANOVA showed that higher level questioning was an

effective instructional strategy in improving writing content from fifth grade students ( $p < 0.05$ ).

Considering all the data on writing presented thus far, in summary, repeated measure ANOVA shows that higher level questioning of students does translate into higher level writing content and does improve students' writing with statistical significance,  $p < 0.05$ . This shows there is now strong evidence to support the use of Bloom's taxonomy as a questioning scaffold in the classroom. While this study measured 6-8 hours of treatment following the oral inquiry format, to improve writing teachers need not spend large amounts of money purchasing a quick fix program or attending a conference to learn a new teaching strategy. Rather, they can improve writing by planning lessons to include using the application, analysis, synthesis and evaluation levels of Bloom's taxonomy as a scaffold to form question stems about the text students will be reading.

Bloom's taxonomy has been in the classroom for years and will continue to be. The writing improvement results will come from the teacher who is willing to spend a small time learning to use the questioning scaffold. All in all, does higher level questioning of students based on Bloom's taxonomy cause writing content to be higher level? The results from this study show with statistical significance, that yes, higher order questioning of text is an effective treatment that can be used to improve students' writing content.

According to the results, the experimental group showed larger gains in holistic writing scores after the participants were instructed using higher order oral inquiry ( $M = 4.55$ ) than the control group after being instructed using lower level questioning ( $M =$

1.95). The interaction between the group assigned and holistic scoring was significant,  $F(1,19) = 18.42, p < 0.05$ .

When investigating the point scores of the participant's writing, the experimental group showed larger gains after the participants were instructed using higher order oral inquiry ( $M = 4.96$ ), than the control group after instructed using lower level questioning ( $M = 2.17$ ). The interaction between the group assigned and point scoring was significant,  $F(1,19) = 13.33, p < 0.05$ .

In contrast to investigating the effects of higher order oral inquiry on writing, question #6 investigates whether teacher questioning from higher levels of Bloom's taxonomy improve reading comprehension more than teacher questioning from lower levels of Bloom's taxonomy. Levene's test of equality of error variances showed that the control and experimental groups did differ in scores of reading comprehension on the DRP and researcher-created tests. Paired t-test showed significant mean increases from the experimental group on the researcher-created test but failed to show significant mean increases on the DRP test scores from either group. According to the repeated measures ANOVA, the results were inconclusive. Descriptive statistics showed improvements in reading comprehension, but the improvements were not statistically significant on either the DRP test or on the researcher-created test.

According to the DRP test scores of reading comprehension, the experimental group showed larger gains after the participants were instructed using higher order oral inquiry ( $M = 30.64$ ) than the control group instructed using lower level questioning ( $M = 24.82$ ). The interaction between the group assigned and the DRP test score was not significant  $F(1,20) = 1.24, p = 0.06$ . According to the results of this study, instruction

using higher order questioning does not significantly improve reading comprehension on the DRP test.

All in all, when investigating the researcher-created test of reading comprehension, the experimental group showed slightly larger gains ( $M = 7.27$ ), than the control group after being instructed using lower level questioning ( $M = 7.09$ ). However, the interaction between the group assigned and the researcher-created test score was not significant,  $F(1,20) = 0.28, p = 0.60$ . According to the results of this study, instruction using higher order questioning does not significantly improve reading comprehension on the researcher-created test.

#### Contributions to Literature

When searching for relevant journal articles in peer reviewed journals, many components of this research were worth investigating. Current research provided limited articles provided information on: guided oral inquiry, scaffolding, higher order questioning, performance based assessment through writing, assessing higher order thinking in writing, and the impact of higher order thinking on reading comprehension and using Bloom's taxonomy as a questioning scaffold. This research investigates all of the above and provides quantitative data to support the use of higher order thinking in the classroom. The scientific community may view this research as an important tool to quantitatively show researchers as well as classroom teachers the data needed to change classroom teaching away from teaching from a purchased self proclaimed, fix all program to an instructional strategy evoking what educators already know coupled with what materials they already have. All in all, this study shows what other researchers and

articles have not, that using an oral inquiry scaffold based on Bloom's taxonomy significantly improves writing in response to reading in 6-8 hours of instruction.

### Practical Implications of the Study

This study offers educators valuable information. This study showed that the improvements seen in students' writing are great. While state and local testing today depends upon the ability of students to demonstrate their understanding of what they've read through writing, using higher order thinking in the classroom, in this study, significantly improves writing in response to reading. Students' writing was higher order, displaying comprehension to include making judgments, suggestions, questioning the author, synthesizing information, forming opinions, analyzing the text and many other skills. Often in classroom today, students write to convey their understanding, but use verbatim duplication of text. Students give the "who, what, when and where," of the story without actually delving into the real meaning the author intended or the meaning of the story as interpreted by the reader. Students become habitualized into writing the sequence of events, paraphrasing, and many other low level skills. As this becomes a habit over time, breaking the habit and requiring more in students' writing, requires different teaching of comprehension and different standards of measurement such as the rubric created and used in this study.

Another complaint about students' writing, comes from critics who say that writing is subjective and not a form of assessment. In this study, the results clearly showed that when two highly qualified scorers assess the same student work, the scores were closely related, compatible and highly reliable. In classrooms today, educators and

parents can rest assured that the scoring of student writing in one classroom can be scored the same as student writing in another classroom.

While some educators are new to holistic scoring preferring to award students points on specifics in their writing. Scoring student writing using the researcher-created rubric, scoring student writing holistically or through a point system are interchangeable. The results of this study tell educators that there isn't a lot of variance in scoring, when using the rubric attached. Whether teachers choose to score writing holistically or by awarding points, this rubric is a good tool for educators to measure higher order thinking in writing. The rubric compliments Bloom's Taxonomy, is easy for educators to use and is the first of its kind, according to a search of recent research. While the direction of scoring students writing as a whole, recently has shifted to include holistic scoring; the higher order questioning in this study, and the rubric-scoring tool, intertwines the belief that higher order thinking as a questioning scaffold, is a necessary teaching strategy. Holistic scoring of writing should be used to measure it.

The improvements of reading comprehension, as seen in this study, are inconclusive. The researcher-created test using question stems with higher order thinking prompts, while difficult to construct was a good measurement of reading comprehension indicating that higher order oral inquiry does improve reading comprehension according to this instrument. The results from the DRP test failed to significantly show that higher order questioning improves reading comprehension perhaps because higher order thinking requires elaborate thought, of which cannot always be answered in a few words in a multiple choice formatted test.

All in all using oral inquiry based on Bloom's Taxonomy has shown to statistically improve students' writing in response to reading. Writing demonstrates higher order understanding of text. Perhaps Wilson (1973) was correct that the taxonomy that the teacher uses influences the level of response among students. Educators should take this into account and instruct students using higher order thinking questioning based on Bloom's Taxonomy.

### Limitations

This study is limited by the small sample size and the creditability of the impact of this study with a small number of participants. DRP is a good test of vocabulary and comprehension although the cloze sentence questions do not necessarily mimic standardized multiple choice formatted test. A better choice for standardized tests of comprehension may have been the Woodcock Johnson Reading Mastery test. Another limitation was in creating a multiple choice test with higher level questioning stems. This was not standardized and to my knowledge, no standardized test exists. Further research is needed to provide additional information.

### Recommendations for Future Research

To begin, further research is needed with a larger sample size to see if the same outcomes are found. With a sample size of 22, it cannot be assumed that this small sample represents the larger population. While the writing of the participants in the treatment or experimental group improved, perhaps with a larger sample size the gains may have greater strength to support the significance. In terms of reading

comprehension, the outcomes from a larger sample size may show not only gains, as seen in the treatment group when tested using the Degrees of Reading Power test, but gains of statistical significance. Perhaps a larger sample size would confirm the statistical gains noted from the treatment group, on the researcher-created test of reading comprehension.

Also, as reported, Bloom's Taxonomy transcends age, subject matter and type of instruction (Hill & McGraw, 1981). Although a review of current research showed several quasi-experimental studies at the college level, future research may want to test Hill and McGraw's statement with elementary and high school participants.

As far as participant writing, although gains were seen it would be interesting to see if writing in response to reading improves across content areas when studied the same. For instance, further research should investigate whether written response in the social studies and science classroom improves when taught using this method. For instance when students are exposed to this scaffold in the language arts class, will writing improvements carry over and be seen in other classrooms as Ge and Land (2001) suggest? Secondly, is it necessary for all content area teachers to teach using this model for gains to be seen in all content area classrooms? Ideally, to improve higher order thinking among students, students need to be taught like this repetitively and students' writing will indicate higher level thinking regardless of the written assessment, whether the assessments are formative, summative, performance based assessment or state testing.

Also with regard to writing and the researched created rubric, further research should investigate how this rubric assesses writing in other areas for different purposes. Again a larger sample size would be helpful to see if the gains noted are consistent.

Furthermore, the experimental group participants in this study were taught for 6-8 hours over 12 days for four weeks using the oral inquiry scaffold. Recommended research would investigate if students should always be taught like this, if intermittent teaching using this scaffold is enough to show writing gains and if the gains in writing in response to reading, although not statistical will be long lasting when this research is finished and beyond. Additionally, what results would be seen if students were taught for an extended amount of time throughout the school year?

The results of this study were inconclusive, but encouraging, about the effects of higher order oral inquiry on reading comprehension. Future research may want to duplicate the format of this study to see if instructing students using higher order questioning for longer than 6-8 hours, improves reading comprehension scores on DRP and other standardized assessments. Perhaps when students are questioned critically, not only do students have a deeper understanding of text, but also inherently, students understand and are able to answer lower level reading comprehension questions. Further investigation is needed before an assumption can be made about the effects of using higher order thinking on reading comprehension.

Furthermore while gains, although not statistical were seen from the treatment or experimental group on the DRP, scores from the researcher-created test were significant. If as Mehrens suggests, performance based assessment is a window dressing and the real stuff is in the standardized form of multiple-choice format (Mehrens, 1992); then another attempt should be made to develop a multiple choice formatted test using higher order thinking stems to see if creating good multiple choice formatted tests with higher order question stems are easily possible.

Additionally, neither the DRP nor the researcher-created tests were modeled after the format used in state testing. Future research should investigate the outcomes of this oral inquiry scaffold on assessments modeled after state tests.

## WORKS CITED

- Albertson, L. and Billingsley, F. (2001). Using Strategy Instruction and Self-Regulation to Improve Gifted Students' Creative Writing. *Journal of Secondary Gifted Education*, 12 (2), 90-102.
- Borsuk, A. (2001). Standardized Assessment is Changing Education. *Milwaukee Journal Sentinel*, June, 17, 2001.
- Benson, M.J. and Sporkowski, M. (1992). Writing Reviews of Family Literature: Guiding Students Using Bloom's Taxonomy of Cognitive Objectives. *Family Relations*, 41(1), 55-70.
- Bloom, B.S., Engelhart, M.D., Furst, F.J., Hill, W.,h., & Krathwohl, D.R. (1956). *Taxonomy of Educational Objectives: Cognitive Domain*. New York: McKay.
- Davey, R. and McBride, S.. (1986). Effects of Question Generation Training on Reading Comprehension. *Journal of Educational Psychology*, 78 (4), 256-262.
- Dewey, J. *The Child and the Curriculum*. Chicago: University of Chicago Press, 1902.
- Dewey, J. *Democracy and Education*. New York: Macmillan Company, 1924.
- Foote, C. (1998). Student Generated Higher Order Questioning As A Study Strategy. *Journal of Educational Research*, 92(2), 107-116.
- Ge, X. and Land, S. (2001). Scaffolding Students' Problem Solving Processes on an Ill-Structured Task Using Questioning Prompts and Peer Interactions. ED 470 086.
- Granello, D. (2000). Encouraging the Cognitive Development of Supervises: Using Bloom's Taxonomy in Supervision. *Counselor Education & Supervision*, 40 (1), 31-47.
- Granello, D. (2000). Contextual Teaching and Learning in Counselor Education. *Counselor Education and Supervision*, 39 (4).
- Graves, D. (2002) *Testing is Not Teaching*, Heinmann: NH
- Guszak, F. (1967) Teacher Questioning and Reading. *The Reading Teacher*, 21 (3), 227-234.

- Guthrie, S. and Davis, M. (2006). Scaffolding for Engagement in Elementary School Reading Instruction. *The Journal of Educational Research*, (Sept-Oct 2006): 3(18).
- Hancock, G. (1994). Cognitive Complexity and the Comparability of Multiple-Choice and Constructed-Response Test Formats. *Journal of Experimental Education*, 62(2), 143-158.
- Hill, P.W., & McGraw, B (1981). Testing the simplest assumption underlying Bloom's Taxonomy. *American Educational Research Journal*. 18, 92-101.
- Hooper, S., Swartz, C., Wakely, M., De Kruif, R., and Montgomery, J.(2002). Executive Functions in Elementary School Children With and Without Problems in Written Expression. *Journal Of Learning Disabilities*, 35 (1).
- Mehrens, W. (1992). article not named in Educational Measurement: Issues and Practice
- Mitchell, R. (1995). The Promise of Performance Assessment: How to Use the Backlash Constructively. Paper presented at the American Educational Research Association annual conference.
- Murphy, N., Messer, D. (2000). Differential Benefits from Scaffolding and Children Working Alone. *Educational Psychology*, 20 (1).
- Pollington, M., Wilcox, B., and T. Morrison (2001). Self Perception in Writing: The Effects of Writing Workshop and Traditional Instruction on Intermediate Grade Student., *Reading Psychology*, 22, 249-265.
- Roid, G.(1994). Patterns of Writing Skills Derived From Cluster Analysis of Direct Writing Assessments. *Applied Measurement in Education*, 7(2), 159-170.
- Ross J., Rolheiser, C., and Hogaboarn-Gray, A. (1998). Effects of Self Evaluation Training on Narrative Writing, *ED 424 248* 1-23.
- Seal, K. (1993) Performance Based Tests: The Reforms Aim to Develop Kids' Thinking and Reasoning Skills, Not Simply Their Ability to Memorize. *Omni*, 16 (3), 66-67.
- Strange, W. (1997). On the Criticisms of Performance Assessment. *Contemporary Education*, 69 (1), 30-34
- Stuhlman, J. Daniel, C., Dellinger, A., Denny, K., and Taylor Powers (1999). A Generalizability Study of the Effects of Training Teachers' Abilities to Rate Children's Writing Using A Rubric. *Journal of Reading Psychology*, 20, 107-127.

- Vogler, Kenneth. (2002). The Impact of High-Stakes, State Mandated Performance Assessment on Teachers' Instructional Practices. *Education, 123*, 39-56.
- Wharton-McDonald, R., Pressley, M., Hampton, J. (1998). Literacy Instruction in Nine First Grade Classrooms: Teacher Characteristics and Student Achievement. *The Elementary School Journal*. (Nov 1998): 101.
- Wilson, I. A. (1973). Changes in the Mean Levels of Thinking in Grades 1-8 Through Use of an Interaction Analysis System Based on Bloom's Taxonomy. *The Journal of Educational Research, 66*(9), 424-429.
- Zemelman, S., Daniels, H. and Hyde, A (1998) Best Practice: New Standards for Teaching and Learning in America's Schools. Heinmann, Portsmouth: NH.
- Samson, G.E., Strykowski, B., Weinstein, T., & Walberg, H. J. (1987). The Effects of Teacher Questioning Levels on Students' Achievement: A Quantitative Synthesis. *Journal of Educational Research, 80*(5), 290-295.

## APPENDIX A

### Higher Order Thinking Rubric

Based on Bloom's Taxonomy

This is used to score writings in two ways. First use this rubric as a holistic scoring piece to give an overall rating on the writing sample. Secondly, this tool can be used to award points to content writing. Hence if one of the criteria listed in the level five section is present in the writing, the student earns five points. Six points for every content presented in the level six section and so forth. An average may be obtained using this method.

<b>6</b>	defends or appraises events criticizes elements in the story debates over information presented in text gives opinions based on contextual knowledge prioritizes information in a hierarchical fashion disputes elements makes an evaluation/judgment
<b>5</b>	formulates a theory about the contextual information proposes alternative events to enhance story line speculates beyond the story or about the events in the story suggests modification interprets the story in a different fashion beyond the literal meaning elaborates on information beyond the literal meaning persuades the reader
<b>4</b>	makes deductions from information in the text discusses cause/effect relationships from story provides evidence to support a statement discusses similarities discusses differences compares to other texts contrasts between texts
<b>3</b>	predicts beyond the story asks questions to illicit more information relates contextual information to other elements in story relates textual information to personal knowledge categorizes information from the reading

<b>2</b>	gives examples outside of text exaggerations generalizes retells through paraphrasing parts of the story orders ideas sequentially
<b>1</b>	gives instructions lists information states who, what, when and/or where information duplicates text verbatim

## APPENDIX B

Researcher-created reading test using higher order thinking questions.

### *The Lucky Cricket* – Pretest

1. If you were to change the animal in the story which animal would also be a good choice?
  - a. firefly
  - b. grasshopper
  - c. praying mantis
  - d. ant
  
2. What word best describes Ling-Ling?
  - a. hopeful
  - b. trusting
  - c. kind
  - d. uncaring
  
3. What might have happened if the cricket hadn't jumped onto Ling- Ling's shoulders?
  - a. the cricket might have gotten away
  - b. the cricket might have spoken up
  - c. Ling-Ling might have seen the snake
  - d. Ling-Ling might have seen another cricket
  
4. What is the motive behind Ling-Ling picking up the cricket?
  - a. she wanted a pet
  - b. to keep from swishing it in the garden
  - c. to keep the cricket quiet
  - d. she believed it would bring her good luck
  
5. What would have resulted if Ling- Ling had been playing in creek?
  - a. she would have found a cricket
  - b. she wouldn't have bad luck
  - c. she wouldn't have found the cricket
  - d. she would have found a grasshopper
  
6. How can you apply this story to your life?
  - a. by watching out more closely for snakes
  - b. by not believing that animals have good luck
  - c. by finding a cricket

- d. by trusting and believing in yourself
7. Why do you think the cricket feels he's unlucky?
- a. because Ling-Ling captured him
  - b. because Ling-Ling can't hear him talking
  - c. because it's dark in her pocket
  - d. because the snake almost bit Ling-Ling
8. Why do you think the snake is important to the story?
- a. to make the cricket realize that he was responsible for saving Ling-Ling's life
  - b. to make the cricket realize he did bring luck
  - c. to make Ling-Ling realize that she is careless
  - d. to make Ling-ling realize that she needs the cricket
9. What was the purpose of the snake, crane and goldfish?
- a. to convince Ling-Ling to believe in the cricket
  - b. to entertain the reader
  - c. to support the belief that the cricket was unlucky
  - d. to support the belief that the cricket stood for good luck
10. What would've happened if Ling-Ling had heard the cricket's thoughts?
- a. the cricket would have been quiet from then on
  - b. the cricket would have convinced he wasn't good luck
  - c. Ling-Ling would have put the cricket down
  - d. Ling-Ling would have cheered the cricket up
11. What would you have added or deleted from the story to make it more interesting?
12. What are your opinions about Ling-Ling capturing the cricket?
13. Would you recommend to this story to a friend? Why or why not?

## APPENDIX C

Researcher-created reading test using higher order thinking questions.

### *Father's New Game* – Posttest

1. What word best describes the father?
  - a. crafty
  - b. creative
  - c. kind
  - d. strict
  
2. What might have resulted if the father hadn't thought of a new?
  - a. Mary and Susan would have created a game
  - b. Mary and Susan would have been made at their father
  - c. Mary and Susan would have watched TV
  - d. Mary and Susan would have fallen asleep
  
3. What is a characteristic of the two sisters?
  - a. hopeful
  - b. patient
  - c. older
  - d. impatient
  
4. Why is it important to know that it was a cold day?
  - a. because this meant the girls couldn't play outside
  - b. because you know snow is probably coming
  - c. because the girls were tired of playing with one another
  - d. because the girls wanted to go outside but weren't allowed
  
5. How did Mary and Susan learn a lesson in the story?
  - a. by not arguing with their father
  - b. by trusting in their father's word and being patient
  - c. by staying in their bedroom while the repair man was there
  - d. by believing in a new game
  
6. What is the father trying to teach the children?
  - a. patience
  - b. honesty
  - c. respect
  - d. fairness

7. How would you determine that Mary and Susan are good kids?
  - a. they were patient while waiting on their father
  - b. they stayed away from the repair man
  - c. they followed their father's directions and kept their promise
  - d. they kept themselves busy by playing together
  
8. What was the importance of the piece of paper on the floor?
  - a. it meant that the girls needed to clean up
  - b. it was the first clue
  - c. it showed the girls that their father was fun
  - d. it showed the girls that the father hadn't forgotten to make a new game
  
9. What information supports the fact that the father was busy?
  - a. he was talking to the repair man
  - b. he was writing all the clues
  - c. he was rushing around the house
  - d. he was waiting for the repair man
  
10. Why was it important for the girls to stay in their room?
  - a. so they won't bother the repair man
  - b. so they won't bother their father
  - c. so they won't see any clues
  - d. so they can learn to trust in what people say
  
11. What would you have added or deleted from the story to make it more interesting?
  
  
  
  
  
  
  
  
  
  
12. What are your opinions about the actions of the father that day?
  
  
  
  
  
  
  
  
  
  
13. Would you recommend to this story to a friend? Why or why not?

APPENDIX D

# Auburn University

Auburn University, Alabama 36849-5212

Curriculum and Teaching  
Haley Center

Telephone (334) 844-4434

**STUDENT ASSENT FORM:**

**A Study of Higher Order Thinking as a Scaffold in the Reading-Writing Connection**

Students,

I would like to learn more about your understanding of a story and how it helps you write. I will be asking your class questions about the story we read and I will record your answers on paper. I will also be asking you to do some drawing, webbing (make a graphic organizer with organized writing and drawing) and writing. This will help me learn more about your reading and writing.

This is called a study. I will be studying how you and your classmates read, answer questions and write about a story.

If you want to be part of this study you have to say it's okay and you have to write your name on the line below. That will let me know it's okay with you.

---

Student's Signature

Thank you for your help with this study.

Ms. Anthony  
Reading Specialist  
Frederick County Public Schools  
Ph.D. Candidate – Reading Education  
Auburn University  
Auburn, Alabama

APPENDIX E

# Auburn University

Auburn University, Alabama 36849-5212

Curriculum and Teaching  
Haley Center

Telephone (334) 844-4434

INFORMED CONSENT FOR:  
**A Study of Higher Order Thinking as a Scaffold in the Reading-Writing Connection**

Your child has been invited, along with his/her classmates, to participate in a study that will involve children in the natural setting of their classroom. This study is being conducted by Ms. Brooke Anthony, Reading Specialist, Frederick County Public Schools and Ph.D. Candidate, under the supervision of Dr. Bruce Murray, Associate Professor in the Department of Curriculum and Teaching, Auburn University. I want to know if students are asked higher-level questions, will their writing improve and will they understand text better? The research doesn't involve any teaching or testing that would be different from normal classroom instruction nor is it an attempt to structure their behavior in any way. This study has been approved by the Frederick County Public School System, Paul Smith, your child's principal, and the Institutional Review Board for the Protection of Human Subjects in Research at Auburn University.

If you decide not to let your child participate in this study, his/her data will not be included. The study will not interfere with students' usual school day or routine. All information will be gathered as the children participate in normal reading activities. Information gathered will be coded so that the participant's identity will be protected. Your child's confidentiality will be protected at all times. My research involves questioning the class, pre and post-test measures to record reading growth and collecting writing samples. Only information that may have bearing on this study will be used.

Information gathered during this study will never be used to identify your child or the school. Results of the study will be reported without using your child's name.

Students may benefit from this study. Students will hear good children's literature, may improve fluency, reading comprehension, decoding skills and may learn more high frequency words. As a result of this study, students may think and write more critically.

Your decision whether or not to allow your child's data to be used in the study will not affect in any way your relationship with Auburn University or the Department of Curriculum and Teaching at Auburn University. Furthermore, your decision will not affect your relationship with Frederick County Public Schools, Ballenger Creek Elementary School or myself, the researcher. You may stop your child's data from being used at any time without penalty or hard feelings. If you decide later that you do not want your child's data to be used in this study, it will be excluded. You may also ask that any information involving your child be destroyed.

If you have any questions now, I invite you to contact me. If you have any questions in the future, please also contact me at (240) 236-4000 or email me at [Brooke.Anthony@fcps.org](mailto:Brooke.Anthony@fcps.org). Additionally my faculty advisor Dr. Bruce Murray can be reached at [murraba@auburn.edu](mailto:murraba@auburn.edu) to answer further questions. For more information regarding your rights as a research participant you may contact the Auburn University Office of Human Subjects Research or the Institutional Review Board by phone (334)-844-5966 or e-mail at [hsubjec@auburn.edu](mailto:hsubjec@auburn.edu) or [IRBChair@auburn.edu](mailto:IRBChair@auburn.edu).

**HAVING READ THE INFORMATION PROVIDED, YOU MUST DECIDE WHETHER OR NOT YOU WISH YOUR CHILD TO PARTICIPATE IN THE RESEARCH PROJECT. YOUR SIGNATURE INDICATES THAT YOU HAVE DECIDED YOUR CHILD MAY PARTICIPATE.**

Participant's Name \_\_\_\_\_

Parent Signature \_\_\_\_\_

Parent's Printed Name \_\_\_\_\_

Date \_\_\_\_\_

Thank you for your consideration of this.

Brooke Anthony  
Reading Specialist  
Frederick County Public Schools  
Ph.D. Candidate – Reading Education  
Auburn University  
Auburn, Alabama