

EFFECTS OF CONCEPTUAL ASSESSMENTS USING TEST DEBATE AND TEST
ANALYSIS ON CRITICAL THINKING SKILLS AND LITERARY ANALYSIS

Nicholas Kowgios

BA English, Lafayette College, 1984
MA Education, Iona College, 1990

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Education in Instructional Leadership

in the

Department of Education and Educational Psychology

at

Western Connecticut State University

2008

EFFECTS OF CONCEPTUAL ASSESSMENTS USING TEST DEBATE AND TEST
ANALYSIS ON CRITICAL THINKING SKILLS AND LITERARY ANALYSIS

Nicholas Kowgios

Western Connecticut State University

Abstract

This study examined the impact of the conceptually designed assessment, test debate and test analysis, on students' critical thinking and ability to analyze literature. The test analysis and test debate process involved three steps: (a) teachers created and administered a multiple-choice exam that asked conceptual questions; (b) students participated in a Socratic test debate in which they were required to support their answers using specific textual references; and (c) students wrote a metacognitive reflection of the evolution of their thought process including an initial interpretation of the question, the points gleaned during the debate, and their final interpretation of the course concept or theme addressed in the question.

Using a sample of convenience ($n = 157$), this study assessed the use of test debate and test analysis in six separate classes among heterogeneously and homogeneously grouped students in grades 6 through 11. The quasi-experimental research design of this study used *The California Measure of Mental Motivation (CM3)*, Advanced Placement English Language and Composition raw scores, and New York State English Language Arts assessments to consider how well the process enhances students' critical thinking skills and students' ability to read and analyze literature.

A two-group and three-group multivariate analysis of variance (MANCOVA) with the Literature Pretest covariate was conducted on the six dependent variables: Literary Analysis, Mental Focus, Learning Orientation, Creative Problem Solving, Cognitive Integrity, and Scholarly Rigor. The data set was analyzed using an independent variable with two levels and three levels.

The two-group MANCOVA data analysis revealed statistically significantly group difference on three of the six dependent variables (Creative Problem Solving, Scholarly Rigor, and Literary Analysis). The three-group MANCOVA produced similar results with regard to significance level, but examination of mean scores was not consistent with the findings of the two-group MANCOVA. A statistically significant effect of the independent variable three groups (trained and treatment, trained no treatment, and no treatment) existed for Mental Focus, Creative Problem Solving, Scholarly Rigor, and Literary Analysis. It can be concluded that the statistically significant multivariate effect was driven in part by the impact of grouping on these dependent variables.

Copyright by

Nicholas Kowgios, Ed.D.

2008

iii




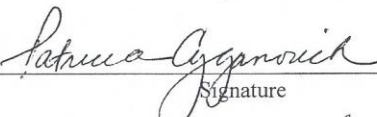

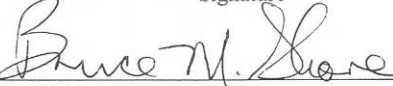
*School of Professional Studies
Department of Education and Educational Psychology
Doctor of Education in Instructional Leadership*

Doctor of Education Dissertation

Effects of Conceptual Assessments Using Test Analysis and Test Debate on
Critical Thinking Skills, Attitudes, and Literary Analysis

Presented by

Nicholas Kowgios

Karen A. Burke, EdD Primary Advisor	 Signature	3/29/08 Date
Patricia Cyganovich, EdD Secondary Advisor	 Signature	3/29/08 Date
Marcia A. B. Delcourt, PhD Secondary Advisor	 Signature	3/29/08 Date
Bruce M. Shore, PhD Secondary Advisor	 Signature	Mar 31, 2008 Date

2008

ACKNOWLEDGEMENTS

My path was smooth because I received impeccable advice. Dr. Karen Burke's tireless efforts to revise and focus my work helped me immeasurably. Her knowledge, work ethic, and understanding motivated me to do my best work. I never wanted to disappoint her, which is the highest praise one can offer. I want to thank Dr. Marcia Delcourt for her wisdom, understanding and foresight as the catalyst of the Doctoral program; and Dr. Bruce Shore for the balance and perspective he provided. Special thanks to Dr. Patricia Cyganovich and Dr. George Bovino for providing me with an opportunity to learn and grow. I am eternally grateful for all the support they have selflessly given me throughout this process and my entire career. I would also like to acknowledge the generosity and support offered by the North Salem Foundation for Learning, a group of individuals with dedication and vision.

DEDICATION

This dissertation is dedicated to my wife, Kathleen, my princess, Taylor, and my best friend, Nicholas. I would accomplish nothing without their love and support. Everyday, they fill my life with joy and meaning.

Table of Contents

	Page
Abstract	i
Copyright	iii
Approval Page	iv
Acknowledgements	v
Dedication	vi
Table of Tables	xi
CHAPTER 1: INTRODUCTION	1
Statement of Purpose	2
Test Debate and Test Analysis Processes Overview	3
Rationale	4
Statement of the Problem	6
Description of Potential Benefits	7
Definitions of Key Terms	8
Research Question	9
Overview of Methodology	9
Chapter Conclusion	10
CHAPTER 2: REVIEW OF THE LITERATURE	12
Theoretical Constructs Underlying Test Analysis and Test Debate	12
John Dewey and Progressive Education	12
Jean Piaget's Genetic Epistemology	14
Vygotsky and Piaget Discord	14

Ausubel's Discovery Learning	18
Historical Relationship to Test Debate and Test Analysis	19
Test Analysis and Test Debate	20
Critical Thinking Research	20
Critical Thinking in Undergraduate Neuroscience	23
Metacognitive Aspect of Critical Thinking	24
Critical Thinking and Gifted and Talented Adolescents	25
Critical Thinking Skills and Accounting Education	26
Critical Thinking Performance and Laboratory Writing	28
Conclusion	30
CHAPTER 3: METHODOLOGY	31
Sample	31
Student Participants	32
Teacher Participants	33
Research Procedure and Design	34
Instrumentation	35
<i>California Measure of Mental motivation (CM3)</i>	36
Literary Analysis Assessments	41
Data Collection	43
Data Analysis	45
Limitations of the Study	46
Statement of Ethics and Confidentiality	46

CHAPTER 4: ANALYSIS OF DATA AND EXPLANATION OF THE FINDINGS	47
Description of the Data	47
Initial Screening Process	47
Code and Value Cleaning	47
Univariate and Multivariate Outliers	49
Descriptive Statistics	58
Analysis of Data	62
Equality of Groups Prior to Treatment	62
Effects of the Two-Groups on the Dependent Variables	64
Two-group Effects Overview	69
Effects of the Three-Groups on the Dependent Variables	70
Three-Group Effects Overview	76
Conclusion	76
CHAPTER 5: SUMMARY AND CONCLUSIONS	78
Summary of the Study	78
Findings	80
Comparison and Contrast of Findings	81
Limitations of the Study	87
Threats to External Validity	88
Threats to Internal Validity	88
Implications	89
Implication of Effects of Two-Group Design	89

Implications of the Effects of Three-Group Design	93
Suggestions for Future Research	94
Qualitative Research and Students' Attitudes	94
Qualitative Research and Teachers' Attitudes	95
Longitudinal Studies On Critical Thinking	96
Studies Focusing On Specific Classroom Activities	96
Summary	97
REFERENCES	98
APPENDIXES	104
Appendix A: Student and Parent/Guardian Consent Form	104
Appendix B: District Permission Form	107
Appendix C: Human Subjects Review Form	109

Table of Tables

		Page
Table 1:	Student Sample	32
Table 2:	Teacher Participants	34
Table 3:	<i>California Measures of Mental Motivation (CM3) Scale Development and Validation Studies</i>	38
Table 4:	Correlations Between the Four Scales of the CM3 and Measures of Student Motivation, Behavior and Achievement	40
Table 5:	Test Analysis / Test Debate Training and Research Schedule	45
Table 6:	Literary Analysis Z Scores	49
Table 7:	Frequencies Statistics	53
Table 8:	Literary Analysis Test Stem-and-Leaf Plot for RGroup	55
Table 9:	Mahalanobis Distance	58
Table 10:	Two-Group Descriptive Statistics	59
Table 11:	Three-Group Descriptive Statistics	60
Table 12:	Two-Group One-way ANOVA	63
Table 13:	Three-Group One-way ANOVA	63
Table 14:	Box's Test of Equality of Covariance Matrices	65
Table 15:	Two-Group Multivariate Tests	66
Table 16:	Levene's Test of Equality of Error Variances	67
Table 17:	Test of Between-Subjects Effects	68
Table 18:	Box's Test of Equality of Covariance Matrices	71
Table 19:	Three-Group Multivariate Tests	72

Table 20: Levene's Test of Equality of Error Variances	73
Table 21: Bartlett's Test of Sphericity	74
Table 22: Tests of Between-Subjects Effects	75

CHAPTER 1

INTRODUCTION

John Dewey (1938) asserted that the conflict that exists between traditional and progressive educational practices drives educators' struggle to balance content focus and conceptual understanding. The traditional view contends that, "The subject matter of education consists of bodies of information and of skills that have been worked out in the past; therefore the chief business of the school is to transmit them to the new generation" (Dewey, 1938, p. 17). Despite nearly 70 years of educational reform, the traditional view still molds educational policy. The practical matter of No Child Left Behind (NCLB) legislation's yearly progress requires teachers to transfer a body of knowledge to students to assure minimum growth and competency. Less grounded in practicality, is the general belief that students should end a school year knowing more facts than when they started.

Since the traditional model presents material in its final form, the information is essentially static. Founded on the assertion that "all principles by themselves are abstract. They become concrete only in the consequences which result from their application" (Dewey 1938, p. 20); the experience of working with academic content gives academic content meaning and promotes true learning. Dewey's progressive view suggested that, "new objects and events be related intellectually to those of earlier experience, and this means that there be some advance made in conscious articulation of facts and ideas" (Dewey, 1938, p. 75).

The theoretical underpinnings of this study are built on Dewey's assertions towards a constructivist model much in the same way Piaget (1969) and Vygotsky (1978) built their theories. Piaget and Vygotsky's shared belief that the formation of intellect is a

process of development facilitated by social experiences creates true understanding. Ausubel (1962) labeled this acquisition of knowledge “meaningful learning,” that is, understanding supported by a connection between new information and existing cognitive structures.

Statement of Purpose

As educational practices evolve, teachers continue to search for ways to facilitate meaningful learning and content retention. The shift in pedagogy designed to promote higher order thinking and critical thinking skills suggests that educational theorists recognize the value of linking content to a conceptual framework, yet today’s educational system is still plagued by inadequate critical thinking on the part of students at every possible level (Norris, 1985). Tsui contended that teachers need to develop pedagogy that fosters critical thinking skills, and researchers must seek “empirical research literature as to specific instructional techniques that effectively enhance students’ abilities to think critically” (2002, p. 741). Since content-laden courses of study favor only short-term retention, the promise of long-term knowledge acquisition might entice teachers to reconsider the basics of content delivery.

The major topic researched was the impact of a conceptually designed course of study in which the assessments serve as learning tools designed to reinforce the major themes of the course. The impact of this approach on student achievement guided the study. More specifically, the study focused on the reflective practices of test debate and test analysis in which students were posed conceptual questions and asked to support contentions using textual information. The three-step process of test debate and test

analysis was studied to determine its impact on students' critical thinking skills and literary analysis.

Test Debate and Test Analysis Processes Overview

The test analysis and test debate processes involved three steps. The first step required teachers to create and administer a multiple-choice examination that asked conceptual questions. These conceptual questions focused on key concepts that were the basis for the unit of study being assessed. The questions had one answer that was the most supported by the classroom instruction and course literature, and four competing answers that were possible, but centrally flawed in some way. The questions focused on a key unit concept or idea that the teacher wished to emphasize at the conclusion of the unit. The second step was a Socratic test debate in which students were required to support their answers using specific textual references. During the debate process students argued for the answer that they felt was the best, most supportable response to the question. During these debates, the teacher posed the question and then allowed students to discuss the answers. The process required teachers to refrain from influencing the path of the debate. At the conclusion of the debate, students voted on the answer that was best supported during the debate. The third step required students to write a metacognitive reflection of the evolution of their thought process. Students were asked to discuss their initial interpretation of the question, and to explain why they made the initial choice during the examination. Students were then asked to discuss the points gleaned during the debate. More specifically, students were required to understand and present the opposing views. The last aspect of these reflections asked students to convey a final

interpretation of the course concept or theme addressed in the question, focusing on the evolution of their thinking from the beginning to the end of the three processes.

Rationale

The importance of this research stems from the foundational belief that learning can only be meaningful if the learner has a conceptual latticework on which to hang the new information learned. Most high achieving students do this instinctively. They have developed an inherent ability to connect what they are learning to what they already know. Thus, the importance for educators is that strategies and educational practices must be employed that require all students to think conceptually, activating existing knowledge when learning new material.

The theoretical research connects test debate and test analysis to the constructs of John Dewey, Jean Piaget, Lev Vygotsky, and David Ausubel. The theoretical foundation of this study is supported by the constructivist view that learning is the process of building knowledge from interpretation of experience. Dewey's (1938) progressive model suggested that information be connected to earlier experiences involving actual life experience and articulation of facts and ideas (Dewey, 1938). Piaget (1969) and Vygotsky (1962) expanded the scope of this constructivist connection by suggesting that knowledge was formed through a process of continuous self-construction, implying that the formation of intellect is a process of development. Ausubel (1962) asserted that learning should activate prior knowledge and make connections during what Ausubel called *discovery learning*. During this process students rearrange information while integrating it with existing cognitive structures

This study's contention was that, as students understand the metacognitive aspects of test debate and test analysis, they greatly improve their understanding of the conceptual aspects of academic courses, and also better master the content acquisition required for any meaningful learning. Some teachers continue to assess students' understanding using traditional methods while others have embraced the progressive view that meaningful learning requires students to maintain an active role in their learning. This exploratory study considered students' perspectives on the cognitive processes associated with test analysis and test debate as a means to determine the effectiveness of this mode of assessment.

Recent research in the area of critical thinking tied the historical perspective to current research by considering how critical thinking provides students with the skills needed to connect course content with true conceptual understanding. Tsui's (2002) case study research contended that today's students are provided with educational experiences that increase content retention, but do not promote students' ability to improve higher order thinking skills. She observed that class discussion and reflective writing, two key components of test debate and test analysis, promote critical thinking.

Lynd-Balta's (2006) study investigated the impact of specific classroom activities on critical thinking skills in an undergraduate neuroscience course. Lynd-Balta concluded that requiring students to apply content knowledge through problem solving promotes critical thinking skills without sacrificing content retention. Using course content to reflect on existing belief systems is similar to the educational goals of test debate and test analysis.

Schwarm and VanDeGrift (2002) studied the impact of classroom assessments on critical thinking by focusing on how classroom assessments can be used to discover students' conceptual understanding as well as content retention. They concluded that as students better recognized their own misconceptions the students were forced to critically think about the depth of their learning through consistent metacognitive processes. Students realized the benefit of thinking critically about the depth of their understanding.

Statement of the Problem

The central issue is that teachers feel restricted by the pressures to teach course content as it is currently assessed on national and state assessments. Because teachers believe that the most important determinate of achievement is retention of content, they spend most of their time focusing on content. The problem arises when students do not understand the significance of the learning so the content becomes disassociated facts. Many educators are not sure how to draw the relationship between content and concepts, operating under the assumption that one only can be presented at the cost of the other.

The problem also stems from the enticing nature of standard assessments of cognitive skills and information retention. Teachers think students are learning if they score well on tests that assess temporary retention of course facts. Because it is easy to evaluate a student's short-term retention of information, teachers latch onto this practice as they struggle to impart knowledge in its final form. A disconnect exists between the appearance of learning, growing out of an artificial suggestion that learning has occurred, and long-term understanding. Students seem to have learned the information at the conclusion of the unit, but the lack of a reflective element to synthesize the learning results in floundering educational systems with sub-par test scores and learning

experiences that do not impact students' existing belief systems. Educators need to find pedagogical methods that are supported both theoretically and through research, and that satisfy educational systems' need to assure content acquisition while promoting the conceptual understanding necessary for long-term retention.

Description of Potential Benefits

This research sought tangible evidence that a course designed with a conceptual framework is the most effective way to construct learning environments. This research examined test debate and test analysis to help determine if the process promotes what David Ausubel (1962) describes as "meaningful learning." The literature suggests that specific pedagogy designed to improve students' critical thinking skills must be developed and evaluated, and this study was designed to do so by exploring the effects of test analysis and test debate on student thinking.

The widespread implementation of a system of assessment that builds content knowledge through an understanding of course concepts could have a significant impact on the way students learn. If teachers administer alternative methods of assessment that enhance students' critical thinking skills while helping students to develop the skills necessary to excel on standardized tests, meaningful educational reform is possible. This research suggests that the process of test debate and test analysis offers students and teachers an assessment method that significantly improves students' performance on standardized assessments while improving students' higher order thinking skills and long-term retention of course concepts and content.

Definition of Key Terms

1. *Course content* is the factual material of any course of study: indisputable dates, historical events, plot elements, or any other objective information that are clearly defined (Nordvall & Braxton, 1996).
2. *Course concepts* are the central premises of any course of study that are accepted by experts in the field. Levels of interpretation may vary, but these concepts are not open-ended although there may be some debate over the veracity of existing theories (Donald, 1983).
3. *Test analysis* is an educational practice with well-defined procedural steps wherein students apply course content when defending a theoretical stance on key literary concepts. Essential parts of test analysis are student debate and reflection.
4. *Traditional strategies* are conventional assessments about which students are asked to recall specific information presented in its final form (Dewey, 1938).
5. *Socratic test debates* are class discussions in which students argue for a particular answer on a conceptual examination using text substantiation and logic to determine either the author's purpose or author's technique (Wellman, 1970).
6. *Meaningful learning* is the process of linking existing knowledge and new information in a specific, tangible way (Ausubel, 1962).
7. *Student behaviors* are the student's observable interactions with classmates and teachers in the form of oral contributions to class discussions about literature (Reynolds, 1946).
8. *Reflective practice* is the process in which students write a three-part test analysis that includes a metacognitive rationale for their original test answer, an analysis

- of the Socratic test debate, and a final assessment of the test question including their current stance on the conceptual answer (Zuber-Skerrit & Fletcher, 2007).
9. *Literary analysis* is a process in which the understanding of a literary work is driven by a focus on interpretation of author's purpose as it is facilitated by author's technique (Walzer, 1991).

Research Question

By using a systematic approach, this research explored the impact of test debate and test analysis on students' thinking through the following question:
What are the identifiable outcomes of test debate and test analysis processes on students' critical thinking skill and literary analysis?

Overview of Methodology

The methodology used in this study was a quasi-experimental design in which a sample of convenience ($n = 157$) was studied to determine the impact of test debate and test analysis on students' critical thinking skills, as measured by *The California Measure of Mental Motivation* (CM3) (Giancarlo & Facione, 2000), and literary analysis, as measured by New York State English Language Arts assessments and Advanced Placement Examinations. In an attempt to insulate the study from teacher dispositions, the research design called for the subdivision of the control group. Thus, data were analyzed using two and three groups. This study assessed the use of test debate and test analysis in six separate classes among heterogeneously and homogeneously grouped students in grades six through eleven. Four of the classes were taught by teachers trained in the process of test analysis and test debate. Two of these classes acted as the treatment group and employed the full intervention ($n = 75$). The four remaining classes were

assessed by traditional means alone, and acted as the control group ($n = 82$). Further data analyses were conducted comparing the sample using three groups: (a) Treatment Group ($n = 75$) composed of students taught by teachers trained in process of test debate and test analysis, and students assessed using test debate and test analysis; (b) Control Group 1 ($n = 45$) composed of students taught by teachers trained in process of test debate and test analysis, and students assessed using traditional assessments; and (c) Control Group Two ($n = 37$) composed of students taught by teachers not trained in process of test debate and test analysis, and students assessed using traditional assessments.

This six-month study included pretests and posttests on literary analysis level using grade specific state and national assessments, and pretests and posttests on critical thinking skills using the CM3 that contained five scales (Mental Focus, Learning Orientation, Creative Problem Solving, Cognitive Integrity, and Scholarly Rigor). Data were analyzed using a one-way multivariate analysis of variance with a covariate (MANCOVA) to determine if the groups were significantly different following the treatment. The analysis of data included a two-group design (treatment and control) to see if the treatment-group students differed significantly from students not using test debate and test analysis, and the subdivided control three-groups design to explore whether or not teacher disposition impacted students' critical thinking skills and literary analysis.

Chapter Conclusion

The central assertion made in Chapter One is that, as teachers continue to search for ways to facilitate long-term retention of both academic content and concepts, the educational system is still plagued by inadequate critical thinking instruction. Educators recognize the need to develop pedagogy designed to promote higher order thinking and

critical thinking skills, yet educational systems often emphasize course content which is most directly assessed on national and state assessments.

The theoretical research suggests that a study of the efficacy of test debate and test analysis will provide educators with insights on how to connect academic content and concepts. The constructivist view that learning is the process of building knowledge from interpretation of experience links historical constructs to recent research in the area of critical thinking, but the paucity of this research suggests there is still a need to establish ways to connect course content with conceptual understanding.

This research sought to examine whether test debate and test analysis help promote what David Ausubel (1962) described as “meaningful learning.” The literature suggests that specific pedagogy designed to improve students’ critical thinking skills must be developed and evaluated, and this study was designed to do so by exploring the effects of test analysis and test debate on student thinking. Extensive implementation of a system of assessment that builds content knowledge through an understanding of course concepts could have a significant impact on the way students learn.

CHAPTER 2

REVIEW OF THE LITERATURE

This review of the literature on test debate and test analysis is presented in two parts. The first part of this discussion, *Theoretical Constructs Underlying Test Analysis and Test Debate*, addresses the theoretical constructs from educational theorists. This historical perspective considers the connection between test debate and test analysis and the theories of John Dewey, Jean Piaget, Lev Vygotsky, and David Ausubel. Particular attention was paid to the relationship between key aspects of these theories and the research topic. The second part of this review of the literature, *Test Analysis and Test Debate*, addresses a more recent appraisal of the literature pertinent to this topic. The main focus of this section was to consider recent research in the related areas of critical thinking.

Theoretical Constructs Underlying Test Analysis and Test Debate

The underlying theoretical foundation of this study is supported by the constructivist view that learning stems from the process of building a base of knowledge from the learner's interpretation of experience. The constructivist umbrella provides a framework for educational theory pertaining to this study of critical thinking and literary analysis as they relate to the test debate and test analysis processes.

John Dewey and Progressive Education

The epistemology of John Dewey (1938), and his support of a progressive form of education, supplied the foundation for constructivist theory. Dewey's basic assumption about the need for a progressive form of education was that "democratic social arrangements promote a better quality of human experience" (Dewey, 1938, p. 34).

Concepts are more effectively assimilated if the process of integration involves a “continuity of experience.” A Socratic discussion capped by a democratic vote by the most invested stakeholders (the students) presents the greatest opportunity for intellectual growth, as long as the discussions that lead to the final student determinations are discriminate. John Dewey noted a central dilemma in educational practices: If traditional classroom pedagogy presents information in a finished form, there is a requirement that, “the future would be much like the past, and yet [course material] is used as educational food in a society where change is the rule, not the exception” (Dewey, 1938, p. 19). The experience of manipulation of course content to support certain assertions about course concepts is the discriminate parameter that leads to meaningful learning. This process requires students to interpret information so that, “every experience both takes up something from those which have gone before and modifies in some way the quality of those which come after” (Dewey, 1938, p. 35).

Thus, Dewey’s progressive model requires that the association of discriminate information be “related intellectually to those of earlier experiences, and this means that there be some advance made in conscious articulation of facts and ideas” (Dewey, 1938, p. 75). In progressive education, the facts and established ideas that are critical to any academic discipline are not abandoned, but manipulated in a way to form long-lasting, meaningful new knowledge. Dewey concluded, that experience, “the means and goals of education” (p. 35), requires that learning must involve actual life experience. In the present study the experience takes the form of test debate and test analysis.

Jean Piaget's Genetic Epistemology

Jean Piaget (1969) embraced the idea that knowledge was formed through a process of continuous self-construction. When children reach the Formal Operational Period (about 11 years onward) they are capable of thinking hypothetically, using abstract reasoning with formally stated propositions such as the conceptual multiple-choice questions used in this study. Piaget's view that conceptual change is an integral part of cognitive development requires teachers to pose conceptual dilemmas that necessitate cognitive conflict and indicate inadequacies in students' thinking. "Cognitive structures change through the processes of adaptation: assimilation and accommodation. Assimilation involves the interpretation of events in terms of existing cognitive structure whereas accommodation refers to changing the cognitive structure to make sense of the environment" (Driscoll, 2000, p. 195).

Through the process of test analysis and test debate, students' thinking indicates the process of assimilation as they weigh existing beliefs gleaned during the reading of text with those of their peers. The test analysis allows students an opportunity to self-regulate their learning reflecting Piaget's notion of accommodation as they amend their understanding using the information acquired during the test debate process. In this way, Piaget's theory is similar in nature to other constructivist perspectives of Lev Vygotsky.

Vygotsky and Piaget Discord

Vygotsky shared Piaget's central belief that it is important to understand the formation of intellect in terms of the process of development but, unlike Piaget, he asserted that no single concept or stage was the static cause. Vygotsky viewed intellectual development as a complex process that unfolded in a variety of ways (Vygotsky, 1962).

While other developmental theorists perceived cognitive development as having a final destination, Vygotsky saw intellectual development as an ongoing process.

The key for Vygotsky was not the end but the beginning (Vygotsky, 1978). Vygotsky used three techniques to examine the origin of intellectual skills. He introduced obstacles that disrupted normal problem solving, he provided external aids to problem solving, and he asked children to solve problems whose demands exceeded their current understanding (Cole & Scribner, 1978). All three steps are significant elements in the test debate and test analysis process. Students are presented with problems in the form of conceptual multiple-choice questions where no single answer is easily dismissed. Instead, meaning is derived by manipulation of course content rather than repetitious recall of previously stated assertions by the teacher. The external aids in this process are the tenets of logical verbal discourse wherein substantiation provides the foundation for an intellectual stance. Lastly, students are challenged to exceed their current understanding through group discourse that involves students verbally debating author's purpose as illustrated by author's technique.

The manipulation of text leads to an enlightening process representative of learning and understanding (Cole & Scribner, 1978). Thus, the process of test debate and test analysis is supported by Vygotsky's premise of mediation in which "in higher forms of human behavior, the individual actively modifies the stimulus situation as part of the process of responding to it" (Cole & Scribner, 1978, p. 14).

Karpov and Bransford (1995) traced the development of learning back to Vygotsky's two types of learning: empirical and theoretical. The empirical model of "comparing a number of different objects, picking out their common observable

characteristics, ... and formulating a general concept about this class of objects” (p. 61) is fostered under a system of traditional school instruction. Theoretical learning, a much more efficient learning process, supplies the student with “general and optimal methods for dealing with certain classes of problems that direct him or her toward essential (not simply common) characteristics of the problems of each class. Then, in the course of specially organized activity, the student masters and internalizes the process of use of these methods” (Karpov & Bransford, 1995, p. 61).

Vygotsky suggested that the acquisition of social experiences through presentation of special psychological tools such as language, mnemonic techniques, formulae, concepts, symbols, and signs leads to theoretical learning (Vygotsky, 1978). The oral exchanges between students as they discuss course concepts allows students to use these tools as they are modeled by more capable learners in the course of this joint activity. “Given to and used by the child first at the external level, these tools then internalize and become the internal possession of the child, altering all his or her mental functions” (Karpov & Bransford, 1995, p. 61). As students work to substantiate a position on a particular course concept, they acquire the psychological tools to promote meaningful learning. Students’ learning progresses from the use of lower level empirical learning to higher level theoretical learning.

The results of comparative analysis of the effectiveness of empirical and theoretical types of learning are quite dramatic. A noteworthy study was carried out by Pantina (1957) on the formation of writing skills in six-year-old children. Two groups were organized: The first group was taught how to write letters of the Russian alphabet using empirical learning, while the second group was taught using theoretical learning.

The empirical group was taught how to write the letters through demonstration and replication. The teacher gave the pattern of the letters to the children, and showed and explained how to write every element of the letter. Students tried to copy this pattern until they were able of performing the task correctly, at which time the pattern of the second letter was given to them, and so on.

The theoretical group's instruction was based on developing the students' attitude toward the analysis of any contour. This method included analyzing the letter to be copied to determine where the direction of the contour changes, placing dots in those positions, and reproducing the same system of dots in another location on the page. The focus of the theoretical group was to view letters as a model of the contour based on the points where the contour changes, and marking those changes with dots.

Both groups learned under the teachers' guidance, but the form of teacher monitoring was significantly different. The empirical group attempted to master concrete writing skills, while the theoretical group tried to understand the concept of contour. The results obtained by the two groups were dramatically different. The empirical group, using trial and error, progressed very slowly requiring approximately 170 attempts to learn to write the first letter. The transfer of skills occurred very slowly with subsequent letters to the optimum number of trials, 20, with the 22nd letter. The total number of attempts needed to write the 22 letters was 1,230 (Pantina, 1957). In contrast, the learning of the theoretical group progressed very quickly. The students needed only 14 trials to successfully write the first letter. They reproduced the process on ensuing letters, mastering the writing of all 22 letters in only 60 attempts (Pantina, 1957). In a similar fashion, the theoretical practice of test debate and test analysis leads students through a

process in which they uncover meaning by constructing a conceptual understanding of author's purpose by considering the author's conceptual focus as it applies to the human condition. Students learn the key points of a work of literature by understanding the relationship of one concept to another, just as the contour dots in the letters are only important in their relationship to the other dots in the letter.

Pantina's (1957) research has been replicated in more recent studies. Researchers consistently conclude that "mechanical drills are much less necessary under theoretical learning, [and] the process of learning becomes interesting for students" (Davydov, 1986). Talyzina concluded that there were far fewer mistakes when learning orientation is guided by the theoretical model in the order of 5% to 6%, and students were more efficient at recognizing mistakes and self-correcting their errors (Talyzina, 1981). As students debated the conceptual basis for an answer, they searched for contextual support and quickly recognized when a flawed assumption guided their initial stance on a problem. Students quickly embraced the pattern of first seeking to construct a conceptual framework for their learning--and the time needed to gain mastery of course content was dramatically reduced. Talyzina suggested that the average reduction was 1.5 to 2%, and sometimes even more (Talyzina, 1981). Thus, theoretical learning resulted in "complete mastery, high level of maintenance, broad transfer, and conscious use by the student" (Karpov & Bransford, 1995, p. 64).

Ausubel's Discovery Learning

Educational psychologist David P. Ausubel (1962) provided a theoretical basis for the notion that learning can acquire greater meaning when the learner activates prior knowledge and makes connections during what he labeled *discovery learning*. As

Ausubel explained, discovery learning requires students to rearrange information while integrating it with existing cognitive structures in a way intended to *create a desired end product*. Ausubel suggested that students derive meaning when they actively interpret experiences while employing internal cognitive operations. Meaningful learning requires one to link potentially meaningful information to existing knowledge in a “nonarbitrary and substantive way” (Ausubel, 1962, p. 213).

The related construct of *potentially meaningful learning* suggests that the learning experience is not something that exists outside the learner, but meaning occurs when learners “actively interpret experiences using certain internal, cognitive operations” (Ausubel, 1962, p. 214). The contention was that the task’s ability to be potentially meaningful becomes an essential condition if learning is to be meaningful. It is important for teachers to pose questions that require application of knowledge gleaned while reading, allowing students to consider how the acquisition of this new information has impacted their existing beliefs.

Historical Relationship to Test Debate and Test Analysis

The historical foundation provided by John Dewey’s (1934) early assertions that democratic social arrangements provide students with opportunities to expand conceptual learning was upheld by Piaget’s (1969) and Vygotsky’s (1978) constructivist stance that when teachers “recognize and honor the human impulse to construct new understandings, unlimited possibilities are created for students” (Brooks & Brooks, 1993, p. 21). Learning environments that promote self-regulated learning lead to what David Ausubel (1962) describes as “meaningful learning” for students, a prerequisite tenet inherent in the test debate and test analysis processes. It is important to now consider recent literature

relatable to the topic of test debate and test analysis as the processes are described on page 2.

Test Analysis and Test Debate

A more recent review of the literature reveals that time has not assuaged the need for pedagogy that promotes critical thinking; the research in the area of critical thinking acquisition remains limited. Since the early 1990s, the wave of interest in critical thinking displays moments of vigor, but critical thinking research has not moved too far from John Dewey's 1930 definition of "searching for material to resolve doubt and dispose of perplexity" (Leopold, 1997, p.1). University of Illinois professor Robert Ennis, a pioneer in critical thinking research, defined critical thinking as "reasonable, reflective thinking that focuses on deciding what to believe and do" (Leopold, 1997, p. 1). The common vein that ties Dewey to current research is the assertion that critical thinking provides students with the skills needed to connect what they learn to existence outside the confines of a conventional classroom setting. Ernest McDaniel, professor of educational psychology at Purdue University explained that "if it's just a piece of isolated knowledge out there, then it's nothing. A mere acquaintanceship with the facts won't impact on the student's belief system. And if no belief system is emerging, no education is taking place" (Leopold, 1997, p. 3).

Critical Thinking Research

Tsui (2002) addressed the issue of critical thinking in her case-study research. Tsui's primary contention is that today's students may be "more highly educated than ever before, [but] they are not necessarily better educated" (Tsui, 2002, p. 740). She based this assertion on the idea that this country's formal education focuses on building

knowledge through content coverage at the expense of skills acquisition. A shift in pedagogy is required where educators focus less on “teaching students what to think [while focusing more on teaching students] how to think” (Tsui, 2002, p. 740).

Qualitative data were collected through a series of site visits to four purposefully chosen case-study institutions between October 1996 and May 1997. A significant portion of the data came from classroom observations and interviews with a minimum of one administrator and five students from each institution. A random sample of 55 individuals was utilized from the 4 sites. The interviews, which on average lasted an hour, were each audiotaped and transcribed verbatim. These semistructured interviews entailed predominantly open-ended questions. Tsui’s definition of critical thinking, borrowed from the manner in which researchers typically operationalize critical thinking (Furedy & Furedy, 1985), provided prior to each interview was: “students abilities to identify issues and assumptions, recognize relationships, make correct inferences, evaluate evidence or authority, and deduce conclusions” (Tsui, 2002, p. 743). Institutions with similar student-body size were selected to insulate the study from differences in critical thinking emanating from size of student population at the four institutions. Each institution had a full-time student population of less than 5,000. Two focus groups were created contingent on students’ SAT scores. The mean SAT score of students at the low selectivity institutions (Schools A and B) was 1000, while that at the high selectivity institutions (Schools C and D) exceeded 1300 (Tsui, 2002). The primary method of data analysis was “explanation building” archived through repeated reviews of all interview transcripts and observational notes constructing meaning from patterns that were identified.

Tsui's findings suggest that those students exposed to writing and revision displayed significantly stronger critical thinking dispositions. As students revised their ideas after input from peers and instructors, the students were inspired "to think more deeply about their own written product and utilize feedback to improve upon it" (Tsui, 2002, p. 747). The test-analysis process utilizes a similar skill bases as students' work to amend their initial stance on the conceptual point posed by the multiple-choice question focusing on author's intent facilitated by author's technique. The written component of the test debate and test analysis processes is an essential element that helps students use "synthesis, analysis and refinement of ideas through the medium of writing" (Tsui, 2002, p. 748).

Tsui also concluded that class discussion appeared to be related to the development of critical thinking skills. Class discussion encouraged students to develop critical thinking skills by providing them a forum to "verbalize and try out ideas" (Tsui, 2002, p. 750). Tsui concluded that class discussion afforded students the opportunity to discuss and disagree with their peers.

The focus of Tsui's study was on the contextual conditions that appear to develop critical thinking skills, and her observations supported the notion that "participation in classroom discussion encourages the exercise of critical thinking skills by allowing students to test out their ideas verbally, to reflect upon the views of one's peers, and to modify critically one's own views through incorporating feedback from others" (p. 754). Tsui's findings supported the idea that "students are more likely to retain ideas when they participate in a dialogue or debate them" (p. 755) and "writing that is likely to be

conducive to critical thinking is that which demands more analysis and less description” (p. 755).

Critical Thinking in Undergraduate Neuroscience

Lynd-Balta (2006) attempted to investigate the impact of specific classroom activities on critical thinking skills in an undergraduate neuroscience course. She broadened the transferability of her investigation by asserting that “critical thinking skills transcend [academic] discipline” (Lynd-Balta, 2006, p. 167), therefore her findings help elucidate the processes inherent in test debate and test analysis. The experimental design used in Lynd-Balata’s study enabled questioning whether or not projects that require students to hone critical thinking skill in the application of course content to the societal concepts relating to the ethical implications of scientific advances, affected students’ examination scores and self- reported perceptions of the course content and concepts. Data were collected during two separate semesters. The treatment group ($n = 18$) used projects to relate course content to course concepts, while the control group ($n = 13$) did not employ the use of projects.

Lynd-Balta found that the exam scores used to determine students’ mastery of course content were not significantly different. The treatment group mean score of 70% (with a range of 50 to 94%), and the control group mean score of 72% (with a range of 54 to 100%) suggested that mastery of content was not affected by critical thinking activities, but students’ perceptions extracted from students’ reflective essays produced data that suggested that the two groups’ perceptions of their own success were significantly different. All members of the treatment group reported that the use of the project had a significant impact on their views but only 42% of the control group

indicated that their perceptions changed. Lynd-Balta concluded that the treatment group exhibited skills that required students to apply content knowledge and problem solving by using higher-level thinking skills without sacrificing content retention. The act of using course content to reflect on existing and emerging belief systems was similar to the educational goals of test debate and test analysis in which students use authors' literary technique to debate and analyze an author's insights on the broader issue of the human condition.

Metacognitive Aspect of Critical Thinking

A study conducted at the University of Washington by Schwarm and VanDeGrift (2002) sought to determine the impact of classroom assessments on the metacognitive aspect of critical thinking. Schwarm and Van De Grift focused on a key element of test debate and test analysis, investigating "how *Classroom Assessment Techniques* (CAT) can be used to discover students' preconceptions and misunderstandings as they learn in the classroom" (p. 2). During two separate academic semesters the researchers gathered data on the sample ($n = 149$) using field notes, surveys, and semistructured interviews with the two instructors. After each class, students were asked to complete a CAT assignment designed to identify students' understanding of course concepts. The assignment asked students to synthesize the lecture by describing the concept presented during that day's lecture in a way "their 10-year-old sister could understand it" (Schwarm & VanDeGrift, 2002, p. 3). The researchers then shared students' responses with the instructors.

The researchers made two assertions resulting from their investigation. The instructors' perceptions of students' understanding changed as the teachers better

understood specific students' difficulties. The second observation drawn was that students came to better recognize their own misconceptions and self-corrected flawed understanding through peer discussion. As the students were forced to critically think about the depth of their learning through consistent metacognitive processes, they realized the benefit of thinking about what they knew and did not know (Schwarm & VanDeGrift, 2002).

Critical Thinking and Gifted and Talented Adolescents

Gunn, Grigg, and Pomahac (2006) studied gifted and talented adolescents' critical thinking. Their study sought to substantiate claims that speed of processing, creativity, and exceptional memory are no longer the earmarks characterizing intelligence. Instead, they asserted, thinking critically about a variety of social, environmental, political and economic problems and issues are integral to intelligence in the 21st century. Gunn, Grigg, and Pomahac came to define critical thinking as an intellectual process of "conceptualizing, applying, analyzing, and synthesizing [which suggests] that critical thinking is a universal process that transcends subject matter" (2006, p. 3). They asserted that these critical thinking dispositions were much more difficult to evaluate than cognitive skills leading to limited research studies performed in the area of critical thinking.

The purpose of the study was to engage a small group of eighth-grade gifted and talented science students using bioethical dilemmas. The 11 gifted and talented students in the study were identified in accordance with the school district's identification plan requiring gifted and talented students to score in the 98th percentile on the *Canadian Cognitive Ability Assessment*. The students were from a mid-sized middle school in a

southern Alberta city. The research team, comprised of two university professors and a classroom teacher, presented and moderated the discussion of bioethical dilemmas. The Cornell Critical Thinking Test (Level X) (2004) was used to monitor students' progress.

Gunn, Grigg, and Pomahac revealed that, if students were engaged in critical thinking activities during this crucial time in their development, critical thinking dispositions could be enhanced. Hence, the students' ability to observe, conceptualize, and analyze improved during the course of the one-year study. In a similar fashion to test debate and test analysis, this research was driven by the infusion of intellectual dilemmas as a pedagogical practice. The results supported the assertion that these practices produce learners who are "self-directed, self disciplined, self-monitored and self-corrective in their thinking" (Gunn, Grigg & Pomahac, 2006, p. 13).

Critical Thinking Skills and Accounting Education

Research by Harris (1998) studied critical thinking skills as they related to the field of accounting education. In the study of four financial accounting classes ($n = 58$), Harris used an analysis of covariance to determine that an increase in metacognitive critical thinking skills correlated with an increase in academic performance. The purpose of this study grew out of the perception by many accounting students that professors provided them with facts without the related conceptual framework for the knowledge.

As the students in Harris's (1998) treatment group used a modified accounting software program designed to increase the learners' metacognitive skills, the treatment group displayed significantly higher motivation as measured using the *Multiple Learning Strategies Questionnaire* (MLSQ). Harris's finding supported the contention that the learning of facts in isolation did not promote meaningful learning for students.

A qualitative study into the critical thinking process of hospitality management students (Thin, 2001) suggested that particular qualities of classroom environment and culture promoted the acquisition of critical thinking skills in students. Using a constant comparative method, data were analyzed revealing 14 themes: acknowledging differences, challenging justifications, norms of leadership, personal ethics, values, beliefs, embedded assumptions, open-mindedness, empathy, personal experience, work experience, subject matter knowledge, and case study. One class of 13 students enrolled in a management course was used as the data source.

Thin found that students used subject-matter knowledge, work experience, personal experience, and case study information to justify decisions, but unlike the process of test debate, rarely quoted these sources as substantiation for their decisions. The students in the study displayed the 14 identified themes designated by the researcher, but a disconnect existed between the students' decisions and the foundation for those decisions. Critical skills theory clearly suggested that the acquisition of critical thinking skills is enhanced by the students' recognition that a warranted stance needs a sound foundation provided by the existing belief system used to draw conclusions. Thin supported this assertion by concluding that "the process of critical thinking [should] be made explicit to students as an enhancement to the problem solving focus of the course" (Thin, 2001, p. 105). As is the case with test debate and test analysis, Thin suggested that students should be encouraged to use additional sources to justify their reasoning.

Lee (2003) studied ways to promote critical thinking skills in emerging learning environments. Using Bloom's (1956) taxonomy of learning objectives, Lee identified critical thinking skills as application, analysis, synthesis and evaluation for her research

sample ($n = 78$). The sample was divided into four groups: one control which received no cognitive structuring method and one group for each of the three different cognitive structuring methods--preparation, cognitive structuring, and role structuring. The transcripts of the posttest and online discussion were analyzed using a series of Chi-square tests to determine group differences on each type of critical thinking skill.

This study produced mixed results that the researcher attributed to the complicated use of technology that decreased the effectiveness of the treatment. There was a disconnect between the seamless connection of the pedagogical practice designed to improve critical thinking and the acquisition of critical thinking dispositions was precipitated by the overriding use of technology. The discontinuity that existed between the discussion session and the posttest session suggested that critical thinking skills improved when the instructional goal are not obfuscated by the extraneous use of technology.

Critical Thinking Performance and Laboratory Writing

Quitadamo and Kurtz (2006) research compared critical thinking performance of students who experienced a laboratory writing treatment with those who experienced traditional quiz-based laboratory in a general education biology course. These researchers used the *California Critical Thinking Skills Test* (CCTST) (Facione, 1991) to determine whether critical thinking performance in the writing group differed significantly from the nonwriting group. The study took place at a state-funded regional comprehensive university in the Pacific Northwest. All participants were nonmajor undergraduates who were taking biology to satisfy their education science requirement. Ten sections of the general education biology sections were included in the one-year study; 4 of the 10

sections implemented a writing component during weekly laboratory meetings ($n = 158$); 6 traditional quiz-based laboratory sections served as a nonwriting control group ($n = 152$).

The writing group significantly improved critical thinking skills whereas the nonwriting group did not. The writing and nonwriting groups were highly similar when they began the term with similar critical thinking ability (45th and 42nd national percentile for writing and nonwriting groups, respectively). By the end of the academic term those in the writing group had improved their critical thinking skill to above the 52nd percentile, while the nonwriting students dropped to below the 40th percentile, Quitadamo and Kurtz concluded that a significant change in critical thinking skills occurred for the writing group, whereas the non-writing group exhibited no change, and that students could considerably improve critical thinking skills within a fairly short nine-week period of time.

Tishman and Andrade (1995) noted that the possible cause for the disconnect between the asserted need for critical teaching in education and meaningful research is the perceived obscure nature of higher order thinking skills and critical thinking borne out of the question: Can critical thinking skills be assessed and taught in a tangible way? They asserted that measures like the CCTDI and the CM3 do identify and measure the dispositions needed for effective critical thinking pedagogy. Thus the educational environment is primed for meaningful research that further studies critical thinking using an educational design.

Conclusion

This review of the literature pertinent to the topic of critical thinking attempted to show how the educational theories of Dewey (1938), Piaget (1969), Vygotsky (1978), and Ausubel (1962) provide the constructs that support the need to teach critical thinking skills. Despite the historical rationale that these theorists provide, a recent review of the literature supports the assertions by Tsui (2002), Gunn, Grigg, & Pomahac (2006) and others that further research is needed to evaluate pedagogical practices designed to promote critical thinking skills and dispositions.

CHAPTER 3

METHODOLOGY

This study was designed to determine the extent to which the process of test debate and test analysis affected a group of middle and high school students' ability to think critically and analyze literature. The research question posed to guide this research in fulfilling the purpose of this study was: What is the effect of test debate and test analysis process on students' critical thinking skill and literary analysis?

This chapter will provide descriptions of participants, the sampling procedures, research design, data-collection procedures, instruments and their reliability and validity, and limitations of the study.

Sample

The participants in this study were a *sample of convenience* selected to suit the purpose of the study. The target population was a group of students in grades 6 through 11. Research was conducted at a small, suburban school district with a total student population of 741 students in grades 6 through 12. Demographically, the student population's socio-economic backgrounds are middle to upper class with a median home income of \$174,000.00 with 2% of students eligible for free lunch. Over 98% of the students are classified as English Proficient. The Student Stability Rate or the percentage of students who also were enrolled in that school at any time during the previous school year was 98%. The Annual Attendance Rate determined by dividing the school's total actual attendance by the total possible attendance is 94% for the years 2002 to 2006. The Student Suspension Rate, determined by dividing the number of students who were suspended from school for one full day or longer anytime during the school year by the

Basic Educational Data System (BEDS) day enrollment, averaged less than 1% for that same time period. The gender breakdown is 54% female and 46% male. There is a small percentage of ethnic diversity that includes 96% white students, 2% Hispanic or Latino students, 1% black students, and 1% Asian or Native.

Student Participants

The target sample is representative of the school population in gender makeup and ethnicity. The sample size of $n = 157$ is outlined in Table 1. The sample was comprised of 43% male students and 57% female with an ethnic diversity consistent with the school population. The ethnic diversity of the actual students participating in this study was 97% white, 2% Hispanic or Latino, and 1% black students.

Table 1

Student Sample

Grade level	Levels of the Independent Variable	Number of Students
6 and 11	Treatment Group:	75
	Teacher trained and	(6th $n= 37$)
	Test analysis and test debate used	(11th $n=38$)
9 and 10	Control (Sub-group A):	44
	Teacher trained and	(10th $n=18$)
	Test analysis and test debate not used	(9th $n=26$)
7 and 9	Control (Sub-group B):	38
	Teacher not trained	(7th $n=19$)
	Test analysis and test debate not used	(9th $n=19$)

Teacher Participants

The six teachers participating in the study had an average of 9.5 years of teaching experience. Four were trained in the implementation of test analysis and test debate during 25 hours of staff-development. Staff development training took place during the month of January 2007 and the research occurred during February, March, April, and May of 2007. Since teachers' dispositions reflected by willingness to train in the process of test analysis and test debate might impact the study, a second control group (made up of two teachers) was used to insulate the study from the effects of teacher dispositions. Table 2 illustrates the characteristics of the teacher participants.

Table 2

Teacher Participants

Teacher	Number of Years Teaching	Number of Years in Current Educational Setting	Independent Variable Grouping	Courses Taught
Teacher A	8	6	Treatment	English 11 AP
Teacher B	10	10	Treatment	English 6
Teacher C	4	2	Control (Sub- group A)	English 9
Teacher D	1	1	Control (Sub- group A)	English 9 Honors
Teacher E	22	20	Control (Sub- group B)	English 10
Teacher F	12	9	Control (Sub- group B)	English 7

Research Procedures and Design

The quasi-experimental design employed in this research used quantitative procedures to investigate the research question using a Pretest-Posttest Non-equivalent Group Design. Each of the six teachers selected two sections of the courses they taught for use in the study. Each teacher was assigned to either the treatment or one of the control groups. Teachers provided all students with a brief explanation of the procedures to be used in the study created by the researcher, and

a permission slip for students' parents or guardians to sign (Appendix A). Only students whose parents or guardians completed the permission slips participated in the study. All other students were given alternative assignments during the data collection process as outlined in the Data Collection section of this chapter.

The categorical independent variable was the class groupings: Group One included the students who received the treatment and used the practice of test analysis and test debate as a reflective practice; Group Two included the students who did not receive the treatment and did not use the practice of test analysis and test debate (Group Two students used traditional testing strategies alone). Group Two was subdivided in an attempt to test the effects of teacher disposition on students' critical thinking skills and literary analysis. Two of the four teachers in Group Two were trained in the process of test analysis and test debate, but those two teachers did not implement the process during the course of the research study, while the remaining two teachers were not trained and conducted their classes without the influence of the training. Analyses of data from these two sub-groups also were interpreted to see if teacher disposition impacted students' learning in comparison to the treatment.

Instrumentation

The instrumentation for this study took two forms. Critical thinking skills were assessed using the *California Measure of Mental Motivation*, and students' ability to analyze literature was gauged using grade-appropriate New York State assessments and Advanced Placement English Language examinations.

California Measure of Mental Motivation

Students' level of critical thinking skills were assessed using the *California Measurement of Mental Motivation* (CM3) (Giancarlo & Facione, 2000). The internal consistency for the 25-item CM3, evaluated using Cronbach's alpha coefficient, was .53 to .83 for the four scales: Learning Orientation, Creative Problem Solving, Mental Focus, and Cognitive Integrity (Giancarlo & Facione, 2000). The CM3 employs a 6-point Likert-type response format, ranging from 1 (Strongly Agree) to 6 (Strongly Disagree). The authors of the instrument contended "Results from four independent and diverse studies demonstrate the suitability of the CM3 as a tool to assess secondary students' disposition toward critical thinking... Scales from the four factors correlated with known measures of student motivation and academic achievement" (Giancarlo, Blohm, & Urdan, 2004, p. 347).

The CCTST is an instrument closely related to the CM3. It is important to consider research that supports the use of the CM3 as a viable tool for assessing critical thinking dispositions. Giancarlo and Facione (2000) researched critical thinking dispositions using the instrument they developed: *The California Critical Thinking Disposition Inventory* (CCTDI). The CCTDI was used to assess college undergraduates, and was a precursor to the *California Measure of Mental Motivation* (CM3) (Giancarlo, Blohm, & Urdan, 2004). Giancarlo and Facione contended that "educators must commit to sharpening students' cognitive skills as well as strengthening their disposition toward critical thinking. Nurturing these opportunities to use thinking to resolve problems inclines students toward doing so" (Giancarlo & Facione, 2000, p. 3). In their article "Assessing Secondary Students' Disposition Toward Critical Thinking: Development of

the California Measure of Mental Motivation,” Giancarlo, Blohm, and Urdan (2004) discussed four independent studies that produced data to support the development of the California Measure of Mental Motivation, or CM3 (2000). Table 3 presents a chronological summary of the sample characteristics and the purposes of the four validation studies. The goals for the studies were to investigate the stability of the four-factor structure of the CM3 through the use of confirmatory factor analysis (CFA) in two independent samples, to examine the range of reliability estimates in terms of internal consistency, and to evaluate evidence for the validity of the tool.

The correlation coefficients presented in Table 4 indicate that all four scales of the CM3 resulted in statistically significant positive correlations with mastery goals, self-efficacy, and self-regulation at the $p < .01$ level (one-tailed). The disposition toward critical thinking was positively related to students’ sense of self-efficacy and of self-regulation in terms of their ability to “modify their behavior and guide their learning experience in the classroom” (Giancarlo, Blohm, & Urdan, 2004, p. 358).

Table 3

California Measures of Mental Motivation (CM3) Scale Development and Validation

	Study Purpose	Design and Sample	Progress in Validation of the CM3
Study 1	Initial pilot study: To gather data on a large, diverse population	Multiple data collection sites across the U.S. $N = 1,378$ Grades 6-12	Narrowed original 100 items to 48 Post-Study 1: 6 items added to 48 items retained for Study 1 analysis
Study 2	Evaluate internal consistency of four CM3 scales: To examine correlations between CM3 and student motivation and academic achievement	Public, coed, northern California $N = 135$ ninth graders; 349 eleventh graders	Further refined 54-item tool Established alpha reliabilities and predictive validity

Table 3 (continued)

California Measures of Mental Motivation (CM3) Scale Development and Validation

	Study Purpose	Design and Sample	Progress in Validation of the CM3
Study 3	Evaluate internal consistency of four CM3 scales: To examine correlations between CM3 and student motivation and academic achievement	Private, all female college preparatory, Missouri $N = 587$ Grades 9 - 12	Replicate factor structure Predictive validity
Study 4	Evaluate internal consistency of four CM3 scales: Evaluate social desirability To examine correlation with motivation	3 Public high schools from 1 school district in northern California $N = 1,008$ Grade 9 - 12	Predictive validity Discriminant validity

(from Giancarlo, Blohm, & Urdan, 2004, pp. 357-358)

Table 4

Correlations Among the Four Scales of the CM3 and Measures of Student Motivation, Behavior and Achievement

	Learning	Creative Problem Solving	Mental Focus	Cognitive Integrity	<i>N</i>
Studies 2 and 4 Results					
Mastery goal	.67**	.46**	.33**	.09*	482
Self-efficacy	.47**	.43**	.34**	.22**	482
Self-regulation (Study 4)	.36**	.31**	.40**	.09*	
SAT9 Math	.18**	.33**	.22**	.25**	379
SAT9 Reading	.13**	.27**	.13**	.43**	387
SAT9 Science	--	.16**	.11*	.22**	380
SAT9 Lang./Writing	.10*	.12**	.09*	.17**	382
SAT9 Social Science	.09*	.13**	--	.18**	379
GPA	.19**	.25**	.35**	.22**	468
Study 3 Results					
PSAT Math	.20**	.37**	.18**	.15**	434
PSAT Verbal	.26**	.31**	.20**	.26**	434
PSAT Writing	.26**	.33**	.21**	.20**	291
PSAT Selection Index	.28**	.40**	.23**	.25**	434
GPA	.40**	.46**	.44**	.21**	580

Note: SAT = Scholastic Aptitude Test; GPA = grade point average; PSAT = Preliminary Scholastic Aptitude Test * $p < .05$, ** $p < .01$. All significance tests were one-tailed. (Giancarlo, Blohm & Urdan, 2004, p. 359).

Four studies supported the CM3 as a viable measure of the disposition toward critical thinking by students in grades 6 to 12. Giancarlo, Blohm, and Urdan (2004) concluded that the CM3 was a valuable self-report indicator, which can be used to evaluate students' disposition toward critical thinking. Although some of the results were modest in nature, the CM3 serves as a viable tool in which to assess the impact of test debate and test analysis on students' critical thinking skills.

Literary Analysis Assessments

This study also used grade-specific New York State English Language Arts assessments and the English Language and Composition Advanced Placement part one raw scores to measure students' ability to analyze literature. Grade 6 and 7 students were assessed using the multiple-choice questions from existing New York State English Language Arts tests. Two separate 26-item tests were used for both the pretest and the posttest. The students' performance was converted into percentage scores for the purpose of comparison. Students in grades 9 and 10 used existing New York State English Regents Examinations to evaluate their literary analysis abilities. A 40-item pretest and posttest used multiple-choice questions from the Task 3 literature section of the exam. The students' performance was converted into a percentage score for the purpose of comparison. Grade 11 students used existing Part 1 sections of the Advanced Placement English Language and Composition Examination to evaluate their literary-analysis ability. The 54-item

pretest and posttest scores were converted to percentage scores for the purpose of comparison.

The three different examinations used to evaluate students' literary analysis ability have established reliability and validity. The New York State Board of Regents English Language Arts examinations have been field-tested and extensive psychometric analyses have been conducted, attesting to the reliability and validity of the test. DeMauro (2001) reported an internal consistency reliability coefficient of .89. Construct validity aligned with New York State English Language Arts standards, evaluated using multilinear regression analyses and post hoc planned quantitative comparisons (Myers, Gamst, & Guarino, 2006). The English Language Arts validity correlations for each grade level exceeded .35. The College Board (2007) asserted that the AP Exams are created, administered, and scored with rigor and attention to statistical standards for reliability and score validity:

To ensure that AP Exams accurately measure college-level knowledge and performance in each discipline, the development process includes college curriculum surveys, pretesting of multiple-choice questions, and college comparability studies. Further, a set number of multiple-choice questions are reused from year to year, making it possible for statisticians to compensate for differences in difficulty between exams of different years. Each exam question is analyzed to ensure that performance on any given question does not greatly vary between set populations, such as males, females, whites, African Americans, and Latinos. On the rare occasions when such analysis shows that the wording of an exam question might have contributed to inequitable performance by one set

population, the question is omitted from the scoring (The College Board, 2007, p. 1).

Data Collection

Data were collected under the guidelines outlined in the Research Procedures and design section. School district approval was sought and granted prior to the collection of data (Appendix B). Sample members were all assigned a six-digit identification number, and each teacher was given a three-digit code to categorize and track students during the duration of the study. When a student was absent from school during the administration of any of the pretests or posttests, that student was eliminated from the research study. All procedures used during data collection were approved by the Human Subjects Research Review Committee at Western Connecticut State University (Appendix C).

During the month of January 2007, four teachers participated in a series of staff development workshops, at which they were provided with a theoretical base for the use of test debate and test analysis, and trained in the process. Training included a complete description of the three-step process and assistance in creating appropriate conceptual multiple-choice questions for use during end-of-unit assessment. The steps in the process were: (a) Teachers created and administered a multiple-choice exam that asked conceptual questions; (b) students participated in a Socratic test debate in which students were required to support their answers using specific textual references; and (c) students wrote a metacognitive reflection of the evolution of their thought process including an initial interpretation of the question, the points gleaned during the debate, and their final interpretation of the course concept or theme addressed in the question. Part of the training process included classroom observations by the four teachers when they

observed the practice of test debate in progress, and follow-up discussions when teachers had the opportunity to read and discuss finished written test analysis by students in the classes they observed.

All participants in the sample were given pretests in both the CM3 and a grade-appropriate literary analysis test during the month of January 2007. Classroom teachers administered the examinations and the completed test booklets were given to the researcher for scoring. The researcher scored the literature tests using a Scantron machine and recorded scores on a Microsoft Excel spreadsheet. The pretest CM3 answer sheets were forwarded to Insight Assessment Inc. for scoring and the results were emailed to the researcher in Microsoft Excel format. During the months of February, March, April, and May the six teachers in the study proceeded to teach literature units in the same way they traditionally taught these topics. At the conclusion of each unit, the two teachers in the treatment group used test debate and test analysis to assess students' learning and the four teachers in the control group assessed students using tests they created.

In June 2007, all members of the sample were given a grade-level appropriate literature posttest which was graded using the same methods as the pretest, and a posttest CM3 that was again forwarded to Insight Assessment Inc. for scoring. Table 5 summarizes the steps in the data-collection phase of the study.

Table 5

Test Analysis / Test Debate Training and Research Schedule

Date	Activity	Duration
January, 2007	Provide Theoretical Foundation	5 Hours
January, 2007	Classroom Observation	5 Hours
January, 2007	Pilot Implementation of Process	10 Hours
January, 2007	Feedback / Reflection / Discussion	5 Hours
January	Initial Data Collection	1 Month
February, March, April, and May 2007	Research Study / Treatment	4 Months
May - June, 2007	Final Data Collection	1 Month

Data Analysis

Data were analyzed using a one-way multivariate analysis of variance with a covariate (MANCOVA). The MANCOVA was used to test the hypothesis that the population means for the dependent variables were the same across all groups (Green & Salkind, 2005). Wilks' lambda was used as the test statistic because of its prevalence in social science literature. Wilks' lambda evaluated the multivariate hypothesis that the population means on the multiple dependent variables were equal across groups (Green & Salkind, 2005).

At the conclusion of the study, statistical analyses were performed using the pretest and posttest scores for all sample members. Data compiled using the CM3 scores and literature analysis tests were analyzed to determine the impact of test debate and test analysis on students' critical thinking skills and literary analysis. The results of this analysis are reported in Chapter 4 of this dissertation.

Limitations of the Study

Because of the complexities of the test analysis and test debate processes, inherent limitations in this study exist. The practical procedures used by each teacher could greatly affect students' critical-thinking skills development. For example, the level to which a teacher actively facilitated students' discussions and interactions could have impacted students' critical thinking and literary analysis. Further study is required to determine to what extent these variables impact students. In an attempt to safeguard against problems associated with teacher dispositions, this study subdivided the control group as outlined in the methodology section.

Statement of Ethics and Confidentiality

Permission to participate in this research was sought from each district's superintendent, each school principal, and all parents of participating students. To assure confidentiality, each participant was assigned a confidential identification number. All data are stored in a locked filing cabinet in the researcher's office and will be maintained there until the findings have been published, accessible only to other researchers for whom the data will prove useful in further comparative analyses and who are enrolled in Western Connecticut State University's Doctor of Education in Instructional Leadership Program.

CHAPTER 4

ANALYSIS OF DATA AND AN EXPLANATION OF THE FINDINGS

The purpose of this study was to test the effects of the specific pedagogical practice called test debate and test analysis on students' critical thinking and literary analysis abilities. The major research question addressed was: What are the identifiable outcomes of test debate and test analysis processes on students' critical thinking skill and literary analysis? The results are presented in four sections: (a) description of the data, (b) the initial screening process, (c) descriptive statistics, and (d) analysis of the findings including tables and text illustrations. Chapter Four illustrates how the findings of the statistical procedures in this study reflect on the research question that guided the investigation.

Description of the Data

The data-analysis section of this dissertation used the results of the CM3 to study the effects of test debate and test analysis on critical thinking skills, and New York State and Advanced Placement assessments to consider the effects of the process on literary analysis abilities of students. The analysis of CM3 data focused on five scales: (a) Mental Focus, (b) Learning Orientation, (c) Creative Problem Solving, (d) Cognitive Integrity, and (e) Scholarly Rigor. Each scale is supported by subscale data. The pretest and posttest data for each of the five scales were analyzed in conjunction with the pretest and posttest literary analysis test scores.

Initial Screening Process

Code and Value Cleaning

The initial data screening process addressed the issue of code and value cleaning.

Once the data set was collected, a verification procedure was followed that checked for the appropriateness of numerical codes for each value in the study (Meyers, Gamst, & Guarino, 2006). The code-cleaning procedures determined whether every value for each case in the study contained only valid numerical codes or values, and if these codes seemed legitimate. The goal of the code cleaning was not to test the veracity of the codes, but to determine if each code was within the specific range.

The first step in the data cleaning involved a simple visual inspection. The data set was examined for missing values. One case in the sample was excluded from the study because of missing values. Under this method of *listwise deletion*, a single missing value on a single variable leads to the elimination of that case from the statistical analysis (Meyers, Gamst, & Guarino, 2006). The sample-size reduction resulting from this listwise deletion of a single case had a minimal impact on the sample size, so this deletion is not expected to increase the estimate of measurement error or drop the n below the level needed for multivariate procedures.

Additionally, a visual examination of the student identification numbers revealed that one case in the study had a mismatched identification number among the data set. The identification number for this student's CM3 posttest was not consistent with the identification numbers used to record other data. This case member also was removed from the sample because of this inconsistency and the lack of a viable way to rectify the mismatched identification numbers. As in the case with multivariate analysis, the large data set precipitated the use of SPSS to provide an efficient means for further data screening (Meyers, Gamst, & Guarino, 2006).

Univariate and Multivariate Outliers

The next step in the code and value-cleaning process involved the detection of univariate and multivariate outliers. Following the recommendation of Hair, Anderson, Tatham, and Black (1998), the values of each variable were converted to standard scores with a mean of 0 and a standard deviation of 1 using SPSS. With a sample size larger than 80 cases, case scores with z scores exceeding ± 3.0 were considered outliers and considered for possible deletion (Hair, Anderson, Tatham, & Black, 1998). This process is represented in Table 6 by the z scores' distribution of the dependent variable Literary Analysis.

Table 6

Literary Analysis Z Scores

Z Score	Frequency	Percent	Valid Percent	Cumulative Percent
-2.03323	2	1.3	1.3	1.3
-1.81341	2	1.3	1.3	2.5
-1.70351	4	2.5	2.5	5.1
-1.59360	3	1.9	1.9	7.0
-1.48369	2	1.3	1.3	8.3
-1.37378	1	.6	.6	8.9
-1.26387	2	1.3	1.3	10.2
-1.17024	2	1.3	1.3	11.5
-1.15396	5	3.2	3.2	14.6

Table 6 (continued)

Literary Analysis Z Scores

Z Score	Frequency	Percent	Valid Percent	Cumulative Percent
-1.08883	3	1.9	1.9	16.6
-1.04405	4	2.5	2.5	19.1
-.93414	3	1.9	1.9	21.0
-.92600	2	1.3	1.3	22.3
-.82423	3	1.9	1.9	24.2
-.76317	4	2.5	2.5	26.8
-.71432	4	2.5	2.5	29.3
-.68176	2	1.3	1.3	30.6
-.60441	4	2.5	2.5	33.1
-.51893	2	1.3	1.3	34.4
-.49450	3	1.9	1.9	36.3
-.44378	2	1.3	1.3	37.6
-.43751	1	.6	.6	38.2
-.38460	2	1.3	1.3	39.5
-.27469	4	2.5	2.5	42.0
-.19327	4	2.5	2.5	44.6
-.11186	3	1.9	1.9	46.5
-.10560	5	3.2	3.2	49.7
-.05487	1	.6	.6	50.3

Table 6 (continued)

Literary Analysis Z Scores

Z Score	Frequency	Percent	Valid Percent	Cumulative Percent
-.03044	2	1.3	1.3	51.6
.05504	2	1.3	1.3	52.9
.06350	2	1.3	1.3	54.1
.13238	1	.6	.6	54.8
.16495	3	1.9	1.9	56.7
.23259	1	.6	.6	57.3
.27486	3	1.9	1.9	59.2
.29521	4	2.5	2.5	61.8
.37663	2	1.3	1.3	63.1
.38477	2	1.3	1.3	64.3
.40168	6	3.8	3.8	68.2
.45804	2	1.3	1.3	69.4
.49468	2	1.3	1.3	70.7
.57077	3	1.9	1.9	72.6
.73986	3	1.9	1.9	74.5
.82441	1	.6	.6	75.2
.86511	1	.6	.6	75.8
.90895	1	.6	.6	76.4

Table 6 (continued)

Literary Analysis Z Scores

Z Score	Frequency	Percent	Valid Percent	Cumulative Percent
.93431	1	.6	.6	77.1
1.02794	1	.6	.6	77.7
1.04422	2	1.3	1.3	79.0
1.07804	6	3.8	3.8	82.8
1.24713	6	3.8	3.8	86.6
1.41622	9	5.7	5.7	92.4
1.48386	1	.6	.6	93.0
1.58532	6	3.8	3.8	96.8
1.59377	1	.6	.6	97.5
1.75441	3	1.9	1.9	99.4
1.92350	1	.6	.6	100.0
Total	157	100.0	100.0	

The six categorical variables (Literary Analysis and the five scales of the CM3) were selected to represent the 35 categorical variables in the data set. Table 7 shows that there were no code violations for these categorical variables. Means and standard deviations on these continuous variables all seemed reasonable. From this initial assessment, it could be concluded that these variables were “clean.”

Table 7

Frequencies Statistics

		Scale 1: Diligent Lit Post- Test	Scale 2: Desire to increase knowledge (posttest)	Scale 3: I/CS: Creative problem solving (posttest)	Scale 4: FM/C: Fair- minded truth seeking (posttest)	Scholarly Rigor (posttest)
<i>N</i> Valid	157	157	157	157	157	157
Missing	0	0	0	0	0	0
Mean	.5625	28.96	32.66	29.15	35.20	27.29
Median	.5500	29.00	32.00	29.00	35.00	28.00
Std. Deviation	.22746	7.884	7.130	7.264	6.951	5.880
Skewness	.024	-.021	-.220	.167	-.415	.172
Std. Error Skewness	.194	.194	.194	.194	.194	.194
Kurtosis	-.956	-.172	.206	.091	.072	-.194
Std. Error Kurtosis	.385	.385	.385	.385	.385	.385
Minimum	.10	6	10	12	10	14
Maximum	1.00	48	50	50	50	43

An assessment of univariate outliers performed using SPSS for the three research groups trained and treatment, trained no treatment, and no treatment, using the same six representative continuous variables Literary Analysis, Mental Focus, Learning Orientation, Creative Problem Solving, Cognitive Integrity, and Scholarly Rigor, produced stem-and-leaf plots for each of the six variables across the three research groups. The representative stem-and-leaf plot for the variable Literary Analysis shown in Table 8 represents the univariate outlier screening performed for the six continuous variables. Gall, Borg, and Gall (1996) contended that the stem-and-leaf display, a condensed graphical presentation of all the individual scores on a particular measure, easily displays the shape and distribution of scores alerting researchers of the need to use statistics that do not assume a normal curve distribution facilitating the detection of outliers.

Table 8

Literary Analysis Test Stem-and-Leaf Plot for R Group trained no treatment

Frequency	Stem	Leaf
2.00	1	00
6.00	1	557777
4.00	2	0002
3.00	2	577
9.00	3	000002222
4.00	3	5557
6.00	4	000222
4.00	4	5577
.00	5	
2.00	5	57
2.00	6	00
2.00	6	57

Stem width: .10

Each leaf: 1 case(s)

Table 8 (continued)

Literary Analysis Test Stem-and-Leaf Plot for R group--trained and treatment

Frequency	Stem	Leaf
2.00	2	99
9.00	3	111558888
6.00	4	004466
16.00	5	0001111333335579
14.00	6	22224455556699
5.00	7	33569
18.00	8	000000444448888888
5.00	9	22266

Stem width: .10

Each leaf: 1 case(s)

Table 8 (continued)

Literary Analysis Test Stem-and-Leaf Plot for R group--no treatment

Frequency	Stem	Leaf
1.00	2	2
2.00	3	77
4.00	4	0256
6.00	5	033377
10.00	6	0122255579
3.00	7	357
5.00	8	00488
6.00	9	022226
1.00	10	00

Stem width: .10

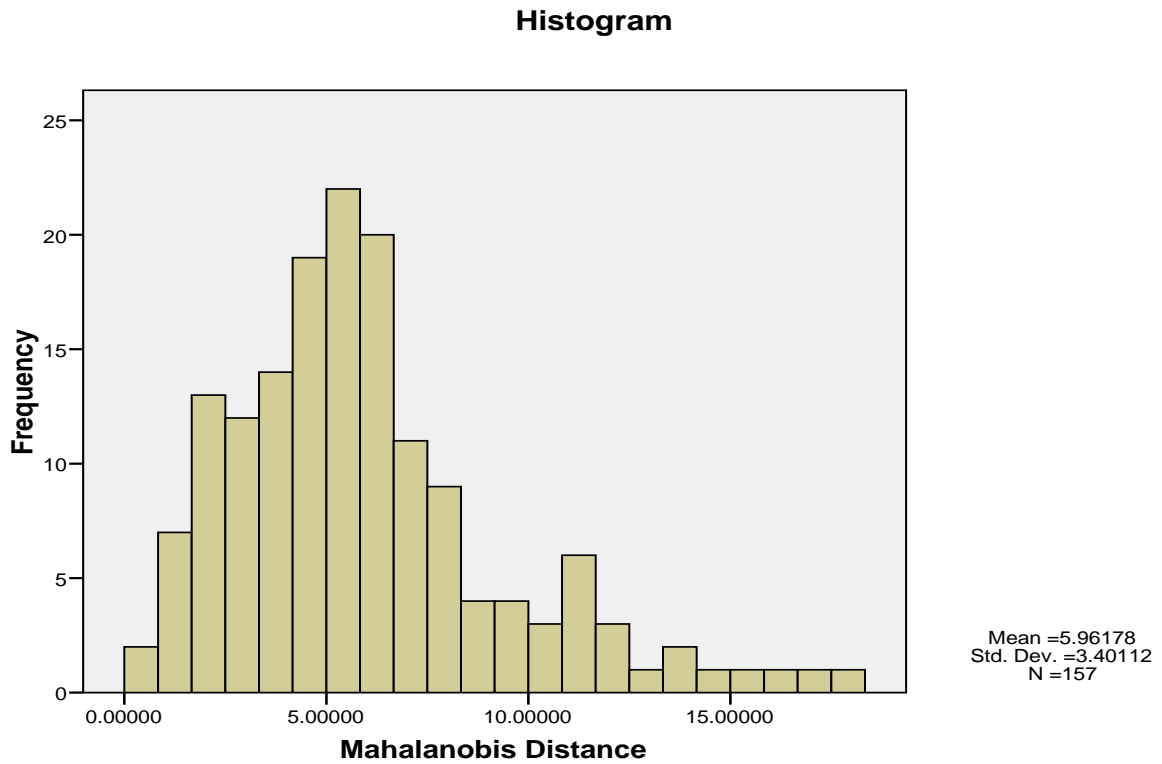
Each leaf: 1 case(s)

This stem-and-leaf plot indicates that no univariate outliers existed for the Literary Analysis variable. Similar results were derived for Mental Focus across all three research groups. Six outliers were present on the remaining four variables across the three research groups, none of which was considered extreme or unusual enough to require deletion. Since none of the scores differed “markedly from the scores obtained by other members of the sample” (Gall, Borg, & Gall, 1996, p. 199), no outliers existed that distorted results.

Following the inspection of the data set for univariate outliers, a consideration of multivariate outliers was needed. Multivariate outliers were screened by computing Mahalanobis distance for each case on the six continuous variables using SPSS. The results detected eight extreme cases at a stringent level of $p > .001$ represented in the Table 9 histogram, none of which was considered extreme or unusual enough to require deletion.

Table 9

Mahalanobis Distance



Descriptive Statistics

The descriptive statistics presented in Table 10 and Table 11 represent the final two-group and three-group data sets used for the statistical analysis following the initial data-screening processes.

Table 10

Two- Group Descriptive Statistics

			Standard	
	Research Group	Mean	Deviation	<i>n</i>
Scale 1: Diligent mental focus	Experimental	30.05	7.175	75
	Control	27.96	8.401	82
	Mean	28.96	7.884	157
Scale 2: Desire to increase knowledge	Experimental	33.84	6.828	75
	Control	31.57	7.269	82
	Mean	32.66	7.130	157
Scale 3: I/CS: Creative problem solving	Experimental	30.27	6.342	75
	Control	28.13	7.917	82
	Mean	29.15	7.264	157
Scale 4: FM/C: Fair- minded truth seeking	Experimental	36.52	7.033	75
	Control	34.00	6.693	82
	Mean	35.20	6.951	157
Scholarly Rigor	Experimental	28.44	5.815	75
	Control	26.24	5.777	82
	Mean	27.29	5.880	157
Literary Analysis	Experimental	.6382	.19233	75
	Control	.4932	.23599	82
	Mean	.5625	.22746	157

Table 11

Three-Group Descriptive Statistics

	Research Group	Mean	Std. Deviation	<i>n</i>
Literary Analysis	Trained and Treatment	.6382	.19233	75
	Trained no treatment	.3415	.14716	44
	No Treatment	.6690	.19401	38
	Total	.5625	.22746	157
Scale 1: Diligent mental focus	Trained and Treatment	30.05	7.175	75
	Trained no treatment	25.95	7.950	44
	No Treatment	30.29	8.408	38
	Total	28.96	7.884	157
Scale 2: Desire to increase knowledge	Trained and Treatment	33.84	6.828	75
	Trained no treatment	30.64	6.878	44
	No Treatment	32.66	7.645	38
	Total	32.66	7.130	157

Table 11 (continued)

Three-Group Descriptive Statistics

	Research Group	Mean	Std. Deviation	<i>n</i>
Scale 3: I/CS:	Trained and Treatment			
Creative problem solving		30.27	6.342	75
	Trained no treatment	26.11	7.406	44
	No Treatment	30.47	7.938	38
	Total	29.15	7.264	157
Scale 4: FM/C: Fair-minded truth seeking	Trained and Treatment	36.52	7.033	75
	Trained no treatment	34.00	6.206	44
	No Treatment	34.00	7.300	38
	Total	35.20	6.951	157
Scholarly Rigor	Trained and Treatment	28.44	5.815	75
	Trained no treatment	24.75	5.392	44
	No Treatment	27.97	5.791	38
	Total	27.29	5.880	157

Analysis of Data

The data analysis sought to determine the effects of test debate and test analysis using quantitative statistical analysis. Initially, a one-way within-subjects ANOVA was conducted to measure the equality of the groups before the study using the Literary Analysis Test Pretest. Second, a two-group between-subjects multivariate analysis of variance (MANCOVA) with the literature pretest covariate was conducted on the six dependent variables: Literary Analysis, Mental Focus, Learning Orientation, Creative Problem Solving, Cognitive Integrity, and Scholarly Rigor. The independent variable was program (treatment versus no treatment). Third, further analysis was performed to determine if the effect of teacher training in the processes of test debate and test analysis--regardless of implementation of the processes--had a statistically significant effect on students' critical-thinking skills and literary analysis. A one-way between-subjects multivariate analysis of variance (MANCOVA) with the literature pretest as covariate was conducted on the six dependent variables. The independent variable was program with three levels, teacher trained and student treatment, teacher trained no student treatment, and teacher not trained and no student treatment.

Equality of Groups Prior to Treatment

A one-way within-subjects ANOVA was conducted to determine if the groups were equal in literary analysis abilities prior to the treatment using both the two-group and three-group model. The results in Tables 12 and 13 revealed that the differences among the literature pretest scores were statistically significant at the $p = .026$ level for the two-group comparison, and at the $p = .000$ level for the three-group comparison. Because the literature pretest scores were statistically significant, comparisons of data

were performed using a MANCOVA. The results indicated that the two-groups and three-groups were different prior to the treatment, which necessitated the use of the Literary Analysis Pretest as a covariate.

Table 12

Two-Group One-way ANOVA

Literary Analysis Pretest

	Sum of		Mean		
	Squares	<i>df</i>	Square	<i>F</i>	<i>p</i>
Between Groups	.281	1	.281	5.058	.026
Within Groups	8.607	155	.056		
Total	8.887	156			

Table 13

Three-Group One-way ANOVA

Literary Analysis Pretest

	Sum of		Mean		
	Squares	<i>df</i>	Square	<i>F</i>	<i>p</i>
Between Groups	1.774	2	.887	19.204	.000
Within Groups	7.113	154	.046		
Total	8.887	156			

Effects of the Two-Groups on the Dependent Variables

A Hotellings's T^2 or two-group MANCOVA was performed on the six dependent variables. The independent variable was program with two levels (treatment versus no treatment). The MANCOVA assessed the effects of one dichotomous (two-group) independent variable (treatment and no treatment) on six quantitative dependent variables: Literary Analysis, Mental Focus, Learning Orientation, Creative Problem Solving, Cognitive Integrity, and Scholarly Rigor using the covariate of Literature Pretest. The use of a MANCOVA provided some control over the alpha level of a Type I error rate, considered dependent variable intercorrelation. This procedure examined the relationships between dependent variables at each level of the independent variable, identified the dependent variables that produced the most group separation, and revealed group differences that may have been masked by univariate analysis (Bray & Maxwell, 1985).

The use of more than one quantitative dependent variable required an examination of the two-group Box's Test of Equality of Covariance Matrices to test homoscedasticity. As evidenced in Table 14, the Box's Test of Equality of Covariance Matrices was not significant (Box's $M = 16.774$, $p = .765$), indicating that the independent variable covariance matrices were equal across the levels of the independent variable. It was appropriate to proceed with the use of multivariate tests since the assumption of homoscedasticity was not violated (Stevens, 2002).

Table 14

Box's Test of Equality of Covariance Matrices

Box's <i>M</i>	16.774
<i>F</i>	.766
<i>df1</i>	21
<i>df2</i>	86878.867
<i>p</i>	.765

An evaluation of the independent variable (group) differences in the population on the dependent variables was determined using Wilks's lambda. Wilks's lambda (Meyers, Gamst, & Guarino, 2006) revealed that there were reliable differences between the treatment and no treatment groups, $F(6, 149) = 18.058, p = .000$ (see Table 15).

Table 15

Two-Group Multivariate Tests

Effect		Value	<i>F</i>	Hypothesis <i>df</i>	Error <i>df</i>	<i>p</i>	Partial Eta Squared
Intercept	Pillai's Trace	.905	236.974	6.000	149.00	.000	.905
	Wilks' Lambda	.095	236.974	6.000	149.00	.000	.905
	Hotelling's Trace	9.543	236.974	6.000	149.00	.000	.905
	Roy's Largest Root	9.543	236.974	6.000	149.00	.000	.905
Literary							
Analysis Test	Pillai's Trace	.673	51.190	6.000	149.00	.000	.673
	Wilks' Lambda	.327	51.190	6.000	149.00	.000	.673
	Hotelling's Trace	2.061	51.190	6.000	149.00	.000	.673
	Roy's Largest Root	2.061	51.190	6.000	149.00	.000	.673
TC							
Group	Pillai's Trace	.421	18.058	6.000	149.00	.000	.421
	Wilks' Lambda	.579	18.058	6.000	149.00	.000	.421
	Hotelling's Trace	.727	18.058	6.000	149.00	.000	.421
	Roy's Largest Root	.727	18.058	6.000	149.00	.000	.421

Because this multivariate test is statistically significant, one can proceed with a separate assessment of each dependent measure (Stevens, 2002). Table 16 illustrates the Levene's Test of Equality of Error Variances, which tests for homogeneity of variance violations for each dependent variable. The evaluation of each dependent measure is not statistically significant ($p < .05$), indicating equal error variance across the two groups.

Table 16

Levene's Test of Equality of Error Variances

	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
Scale 1: Diligent mental focus	.724	1	155	.396
Scale 2: Desire to increase knowledge	.653	1	155	.420
Scale 3: I/CS: problem solving	2.857	1	155	.093
Scale 4: FM/C: Fair-minded truth seeking	.843	1	155	.360
Scholarly Rigor	.029	1	155	.865
Literary Analysis	1.971	1	155	.162

Each dependent variable was evaluated separately in the Tests of Between-Subjects Effects (see Table 17). Statistically significant effects of the independent variable, program (two groups of treatment and no treatment), for Creative Problem Solving, Scholarly Rigor, and the Literary Analysis, suggests that the significant multivariate effect was driven in part by the impact of two levels of the program on Creative Problem Solving, Scholarly Rigor, and Literary Analysis.

Table 17

Test of Between-Subjects Effects

Source	Dependent Variable	Type III		Mean Square	<i>F</i>	<i>p</i>	Partial
		Sum of Squares	<i>df</i>				Eta Squared
Corrected Model	Scale 1: Diligent mental focus	171.098(a)	2	85.549	1.383	.254	.018
	Scale 2: Desire to increase knowledge	221.470(b)	2	110.735	2.212	.113	.028
	Scale 3: I/CS: Creative problem solving	347.592(c)	2	173.796	3.394	.036	.042
	Scale 4: FM/C: Fair-minded truth seeking	274.383(d)	2	137.192	2.909	.058	.036
	Scholarly Rigor	217.574(e)	2	108.787	3.236	.042	.040
	Literary Analysis	5.667(f)	2	2.833	181.49	.000	.702

Group means for each dependent variable (Table 10) revealed that the treatment group had statistically higher scores on Creative Problem Solving ($M = 30.27$, $SE = 6.342$) than did the no-treatment group ($M = 28.13$, $SE = 7.917$), the treatment group had

higher scores on Scholarly Rigor ($M = 28.44$, $SE = 5.815$) than did the no-treatment group ($M = 26.24$, $SE = 5.777$), and the treatment group had higher scores on the Literary Analysis ($M = .6382$, $SE = .19233$) than the no-treatment group ($M = .4932$, $SE = .22746$).

Two-Group Effect Overview

A Hotelling's T^2 or two-group between-subjects multivariate analysis of variance with a covariate (MANCOVA) was conducted on the six dependent variables: Literary Analysis, Mental Focus, Learning Orientation, Creative Problem Solving, Cognitive Integrity, and Scholarly Rigor. The independent variable was program (treatment and no treatment). The covariate was the Literary Analysis Pretest.

Using the Wilks's criterion, three of the six dependent variables were significantly affected by group. Means and standard deviations of the six dependent variables for the two groups were examined. The treatment group achieved significantly higher scores on three of the dependent variables: Literary Analysis, Creative Problem Solving and Scholarly Rigor. Although an examination of mean scores reflected improvement, no statistically significant group effects were observed for Mental Focus, Learning Orientation, and Cognitive Integrity. The implications of these findings will be discussed in Chapter 5.

Effects of Three-Groups on the Dependent Variables

In an attempt to assess the effects of the independent variable (program) with three levels (teacher trained and treatment, teacher trained no treatment, teacher not trained and no treatment) on the six dependent variables (Literary Analysis, Mental Focus, Learning Orientation, Creative Problem Solving, Cognitive Integrity, and Scholarly Rigor) a MANCOVA was applied with the covariate Literary Analysis Pretest. The no-treatment group used in the two-group model was subdivided in an attempt to test the effects of teacher disposition on students' critical-thinking skills and literary analysis. Two of the four teachers in the no-treatment group were trained in the processes of test analysis and test debate, but those two teachers did not implement the processes during the course of the research study, while the remaining two teachers were not trained and conducted their classes without the influence of the training. Analyses of data from these two subgroups were also interpreted to see if teacher disposition impacted students' learning in relation to the treatment.

A one-way within-subjects ANOVA was conducted to determine if the groups were equal in literary analysis abilities prior to the treatment using the three-group model. The differences in the literature pretest scores were statistically significant at $p = .000$. Because the Literary Analysis Pretest score differences were statistically significant, comparisons of data were performed using a MANCOVA. As in the two-group analysis, the three-groups were different prior to the treatment, which necessitated the use of the Literary Analysis Pretest as a covariate.

As was also the case with the two-group MANCOVA, the use of more than one quantitative dependent variable required an examination of the Box's M Test of Equality

of Covariance Matrices to test homoscedasticity. As shown in Table 18, the Box's Test of Equality of Covariance Matrices was not significant (Box's $M = 31.756$, $p = .921$). It was appropriate to proceed with the use of multivariate tests since homoscedasticity was not violated (Stevens, 2002).

Table 18

Box's Test of Equality of Covariance Matrices

Box's M	31.756
F	.709
$df1$	42
$df2$	43152.016
p	.921

To test for treatment effects the next step was to test if a statistically significant multivariate effect was present suggesting that the independent variable (program) was associated with differences between the vectors or sets of means. An evaluation of the independent variable (program) differences in the population on the dependent variables was determined using the Wilks's lambda, revealing there were reliable differences between the treatment and no treatment groups, $F(12, 296) = 11.93$, $p = .000$ (see Table 19).

Table 19

Three-Group Multivariate Tests

Effect		Value	<i>F</i>	<i>df</i>	Error <i>df</i>	<i>p</i>	Partial Eta Squared
Intercept	Pillai's Trace	.892	203.987	6.000	148.00	.000	.892
	Wilks' Lambda	.108	203.987	6.000	148.00	.000	.892
	Hotelling's Trace	8.270	203.987	6.000	148.00	.000	.892
	Roy's Largest Root	8.270	203.987	6.000	148.00	.000	.892
Literary Analysis	Pillai's Trace	.607	38.048	6.000	148.00	.000	.607
	Wilks' Lambda	.393	38.048	6.000	148.00	.000	.607
	Hotelling's Trace	1.542	38.048	6.000	148.00	.000	.607
	Roy's Largest Root	1.542	38.048	6.000	148.00	.000	.607
RGroup	Pillai's Trace	.564	9.761	12.000	298.000	.000	.282
	Wilks' Lambda	.454	11.93	12.000	296.000	.000	.326
	Hotelling's Trace	1.162	14.233	12.000	294.000	.000	.367
	Roy's Largest Root	1.126	27.960	6.000	149.000	.000	.530

Separate Levene's tests for each dependent variable were also not statistically significant, indicating equal variances for each dependent measure across the levels of the groups (Table 20).

Table 20

Levene's Test of Equality of Error Variances

	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
Literary Analysis	2.183	2	154	.116
Scale 1: Diligent mental focus	.289	2	154	.749
Scale 2: Desire to increase knowledge	.476	2	154	.622
Scale 3: I/CS: Creative problem solving	1.059	2	154	.349
Scale 4: FM/C: Fair-minded truth seeking	.759	2	154	.470
Scholarly Rigor	.425	2	154	.654

A statistically significant Barlett's test of sphericity ($p < .001$) as shown in Table 21 indicated sufficient correlation between the dependent variables to proceed with the analysis.

Table 21

Bartlett's Test of Sphericity

Likelihood Ratio	.000
Approx. Chi-Square	1376.164
<i>df</i>	20
<i>P</i>	.000

Each dependent variable was evaluated separately in the Tests of Between-Subjects Effects (Table 22). A statistically significant effect of the independent variable, program (trained and treatment, trained no treatment, and no treatment) existed for Mental Focus, Creative Problem Solving, Scholarly Rigor, and Literary Analysis. Thus it can be concluded that the statistically significant multivariate effect was driven in part by the impact of differences between levels of the independent variable on these dependent variables.

Table 22

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III					<i>p</i>	Partial Eta Squared
		Sum of Squares	<i>d</i> <i>f</i>	Mean Square	<i>F</i>			
Corrected Model	Literary Analysis	6.049	3	2.016	152.582	.000	.749	
	Scale 1: Diligent mental focus	633.393	3	211.131	3.564	.016	.065	
	Scale 2: Desire to increase knowledge	285.189	3	95.063	1.902	.132	.036	
	Scale 3: I/CS: Creative problem solving	593.830	3	197.943	3.965	.009	.072	
	Scale 4: FM/C: Fair- minded truth seeking	279.762	3	93.254	1.966	.121	.037	
	Scholarly Rigor	401.429	3	133.810	4.100	.008	.074	

An examination of group means for each dependent variable (Table 10) reveals that the treatment group had higher scores on Mental Focus ($M = 30.05$, $SE = 7.175$) than did the trained no-treatment group ($M = 25.95$, $SE = 7.950$), but not significantly greater than the no-treatment group ($M = 30.29$, $SE = 8.408$); the treatment group had higher scores on Creative Problem Solving ($M = 30.27$, $SE = 6.342$) than did the trained no-treatment group ($M = 26.11$, $SE = 74.06$) but not greater than the no-treatment group ($M = 30.47$,

$SE = 7.938$); the treatment group had higher scores on Scholarly Rigor ($M = 28.44$, $SE = 5.815$) than did the trained no-treatment group ($M = 24.75$, $SE = 5.392$) but not the no-treatment group ($M = 27.97$, $SE = 5.791$); and the treatment group had higher scores on Literary Analysis ($M = .6382$, $SE = .19233$) than did the trained no-treatment group ($M = .3415$, $SE = .14716$) but not the no-treatment group ($M = .6690$, $SE = .19401$).

Three-Group Effects Overview

A k-group or multivariate analysis of variance with a covariate (MANCOVA) was conducted on the six dependent variables: Literary Analysis, Mental Focus, Learning Orientation, Creative Problem Solving, Cognitive Integrity, and Scholarly Rigor. The independent variable was the program with three levels (trained and treatment, trained no treatment, no treatment). The covariate was the Literary Analysis Pretest.

Using Wilks's criterion, four of the six dependent variables were significantly affected by group. Means and standard deviations of the six dependent variables for the three groups were examined. The treatment group achieved significantly higher scores on four of the dependent variables: Literary Analysis, Mental Focus, Creative Problem Solving and Scholarly Rigor. An examination of mean scores did reflect that statistically significant group effects were observed for the four dependent variables when comparing the treatment group with the trained no-treatment group, but that effect was not present when comparing the treatment group with the no-treatment group.

Conclusion

A Hotellings T^2 or two-group between subjects multivariate analysis of variance with a covariate (MANCOVA), and a k-group or multivariate analysis of variance with a covariate (MANCOVA) was conducted on the six dependent variables: Literary

Analysis, Mental Focus, Learning Orientation, Creative Problem Solving, Cognitive Integrity, and Scholarly Rigor. The independent variable was program with two levels (treatment, no treatment), and three levels (trained and treatment, trained no treatment, no treatment). The covariate was the Literary Analysis Pretest. For both the two-group and three-group MANCOVAs there were significant statistical differences between and among the groups. The implications of these findings will be discussed in Chapter Five.

CHAPTER 5

SUMMARY AND CONCLUSIONS

The six sections of Chapter Five elaborate on the central premise of this study through discussion, description, and extension. The Summary of the Study gives an overview of the entire study, and the Findings section provides a review of the findings from the statistical analysis of test debate and test analysis as they relate to the literature review in Chapter Two. The Comparison and Contrast of Findings section analyzes, evaluates, and synthesizes the research findings of the MANCOVA analyses performed using the two-group and three-group designs. The Limitations section expands on assertions made in Chapter Three through a practical look at specific issues raised during the research study. The Implications section provides suggestions for what should be done as a result of this study of test debate and test analysis, and discusses how to execute these suggestions. Lastly, the Future Research section offers suggestions on what should be done to further the research in the area of critical thinking education.

Summary of the Study

The problem that precipitated the need for this study of test debate and test analysis stems from the recent pressures imposed on educational systems. The influx of standardized assessments as measures used to determine the efficacy of schools has led to pedagogical practices that stress surface-level content retention over meaningful learning. Teachers work to assure content retention at the cost of conceptual understanding. Educators extol the values of critical thinking as a worthwhile educational pursuit without developing classroom practices that promote critical thinking. This study sought to test the effects of a specific educational practice, test debate and test analysis,

through a quantitative analysis of students' critical-thinking skills and students' ability to analyze literature. The research question that guided the research was: What are the identifiable outcomes of test debate and test analysis processes on students' critical thinking skill and literary analysis?

Data were collected in two forms: (a) Critical-thinking skills were evaluated using *The California Measure of Mental Motivation* (Giancarlo & Facione, 2000), and (b) literary analysis skills were assessed using national and New York State standardized tests. A two-group MANCOVA and a three-group MANCOVA were performed on the six dependent variables: Literary Analysis, Mental Focus, Learning Orientation, Creative Problem Solving, Cognitive Integrity, and Scholarly Rigor using the Literature Pretest as a covariate. The quasi-experimental research design employed in this research used quantitative procedures to investigate the research question using a Pretest-Posttest Non-equivalent Group Design.

Six teachers participated in the study. Each of the six teachers selected two sections of the courses they taught. These sections served as the basis for the independent variable, program. The design tested the effects of test debate and test analysis by separating the sample ($n = 157$) into two groups (treatment, and no-treatment). The effects of teachers' disposition in the research study were further analyzed by dividing the "no-treatment" group into two additional levels (teacher trained no-treatment, and teacher not trained no-treatment).

The participants in this study were a sample of convenience. The target population was a group of students in grades 6 through 11 ($n = 157$). Research was

conducted at a small, suburban school district, and the target sample is representative of the school population in sex and ethnicity.

Findings

A two-group and three-group multivariate analysis of variance (MANCOVA) with the Literature Pretest covariate was conducted on the six dependent variables: Literary Analysis, Mental Focus, Learning Orientation, Creative Problem Solving, Cognitive Integrity, and Scholarly Rigor. The data set was analyzed twice using an independent variable with two levels and again with three levels.

In the two-group MANCOVA, the groups were significantly different on three of the six dependent variables (Creative Problem Solving, Scholarly Rigor, and Literary Analysis) at the $p < .05$ confidence level. The partial Eta-squared effect sizes for two of the three dependent variables (Creative Problem Solving and Scholarly Rigor) were modest 3.6% and 4.2% levels, respectively, while an effect percentage of 70% was noted for the Literary Analysis variable. Means scores for the experimental group were higher on all six dependent variables.

The three-group MANCOVA produced similar levels of significance, but applied to different variables as compared to the findings for the two-group MANCOVA. Statistically significant differences on the independent variable for three groups (trained and treatment, teacher trained no-treatment, teacher not trained and no-treatment) existed for Mental Focus, Creative Problem Solving, Scholarly Rigor, and Literary Analysis. Thus, the statistically significant multivariate effect was driven in part by the impact of group differences on these dependent variables. The partial Eta-squared effect sizes for the four dependent variables showing significant difference were 6.5% for Mental Focus,

7.2% for Creative Problem Solving, 7.4% for Scholarly Rigor, and 74.9% for the Literary Analysis. In both the two- and three-variable analyses, Literary Analysis reflected the largest effect size.

Statistically significant group mean differences were not observed for the four dependent variables when comparing the teacher trained no-treatment group with the teacher not trained and no-treatment group. These findings will be discussed in the implications section of this chapter.

Comparison and Contrast of Findings

The Review of the Literature presented in Chapter Two suggested that critical thinking research has established historical roots growing out of the constructs of John Dewey, Jean Piaget, Lev Vygotsky, and David Ausubel. Despite the depth and related nature of these historical theorists, a paucity of meaningful research has existed, testing pedagogy that supports students' acquisition of critical thinking skills (Tsui, 2002). The disconnect between theory and practice supported the need for this research study on the effects of test debate and test analysis. This study supported the assertion that a course of study guided by a conceptual connection between the course content and the course concepts promotes meaningful learning that affects the ways students think critically and analyze literature. Student participation in the metacognitive practice of test debate and test analysis displayed significant improvement in the way they think critically, which translated as well into better content retention. Focus on meaningful learning is not at the expense of factual or content learning, but focus on the latter does not assure conceptual understanding.

John Dewey's (1938) assertions about the need for progressive educational practices are ironically still relevant. This 70-year-old philosophy seems as pertinent today as it was when it was first espoused. If traditional educational practices present information in a finished form, there is a requirement that, "the future would be much like the past, and yet [course material] is used as educational food in a society where change is the rule, not the exception" (Dewey, 1938, p. 19). This study supported the assertion that innovative pedagogy, like test debate and test analysis, needs to be developed if educational systems hope to develop students as thinkers capable of resolving cognitive dilemmas imposed by the changes of the future.

Today's educational leaders constantly stress the need for pedagogy that prepares students to adapt to the ever-changing global community, yet specific classroom methods that promote critical thinking skills remain scarce. As Thomas L. Friedman (2005) explained in *The World Is Flat*, "Our ability to get by doing things the way we've been doing them ... will not suffice anymore" (p. 361). Most educators agree that students need to be critical thinkers capable of resolving complex problems that we have never faced before. Dewey's advocacy for an educational system that nurtures thinking as a primary educational goal challenges teachers to consider what they value by assessing what is important. The literature is clear: Teach students to think through and students will be prepared to face the challenges of the future.

As Dewey explained "democratic social arrangements promote a better quality of human experience" (Dewey, 1938, p. 34). The processes of test debate and test analysis send a clear message to students: There is value in the exchange of ideas. Students learn to incorporate the insight of others into their existing belief systems, formulating refined

conceptual understanding of course material. Course content and concepts become “related intellectually to those of earlier experiences, and this means that there be some advance made in conscious articulation of facts and ideas” (Dewey, 1938, p. 75). Test debate and test analysis helps students “learn how to learn” by approaching learning with “passion and curiosity” (Friedman, 2005, p. 302). This study revealed that, as students manipulate the content that is part of any course of study, they expand their existing belief systems, incorporating new information with what they already understood to be true. This pedagogical practice is a model worth emulating as teachers strive to enhance students’ understanding and long-term retention of course material.

Jean Piaget’s (1969) idea that continuous self-construction promotes cognitive development is a foundational assumption driving educational reform. If students are expected to think critically, they must be given opportunities to do so. As teachers pose conceptual dilemmas that necessitate cognitive conflict in the form of conceptually based questions, inadequacies in students’ thinking become apparent to students. Students learn how to monitor their thinking metacognitively constructing an understanding that is so much more meaningful than the static exchange of facts, which predominantly drives traditional education. The more students work with ideas, the more cognitive structures change. This study supports the contention that as students learn the skills necessary to resolve complex problems through the processes of adaptation and assimilation (Driscoll, 2000), they begin to develop a skill base that can be applied when faced with cognitive predicaments in the future.

The processes of test debate and test analysis are supported by Vygotsky’s (1962) premise that students must exceed their current understanding. Test debate and test

analysis encourages teachers to provide students with learning opportunities in which “the individual actively modifies the stimulus situation as part of the process of responding to it” (Cole & Scribner, 1978, p. 14). Ausubel (1962) concisely captured the notion of learning that requires critical-thinking with his construct of meaningful learning. Meaningful learning requires students to link potentially meaningful information to existing knowledge in a “nonarbitrary and substantive way” (Ausubel, 1962, p. 213). In this way, the learning experience is not something that exists outside the learner, but meaning occurs when learners formulate linked conceptual understanding. The act of debating a particular conceptual stance about author’s purpose in a literary work impacted students’ critical thinking in a way that promoted a dispositional change. The statistically significant differences in students’ critical thinking skills reflected in the data set suggest that test debate and test analysis provided students with the type of meaningful educational experience described by Ausubel.

Recent research in the area of critical-thinking depicts the continued need to construct learning environments that afford students the chance to think, reflect and interact, staples of effective critical-thinking pedagogy. Tsui’s (2002) research suggested that a shift in pedagogy is required wherein educators focus less on “teaching students what to think [while focusing more on teaching students] how to think” (p. 740). Test debate and test analysis are processes that encourage students to think rather than merely regurgitate previously stated ideas. As students debate, they soon realize that the right answer often hinges on substantiated perspective. As students think more deeply about the conceptual underpinnings of course material, they confidently engage in metacognitive reflection as part of the learning process. Tsui’s research identified factors

that appear to be related to the development of critical thinking skills. Class discussion encouraged students to develop critical thinking skills by providing them a forum to “verbalize and try out ideas” (Tsui, 2002, p. 750).

As students work with ideas and debate the merits of any ideological stance they come to value the power of thinking over the memorization of facts. One of the primary attributes of test debate and test analysis supports Tsui’s contentions; test debate and test analysis provides students with a forum to construct meaning through oral discourse. Tsui’s research targeted the specific classroom practice of class discussion as a way to improve critical thinking skills, but fell short of identifying the aspects of such discussions that provide cognitive growth. Tsui’s qualitative observations provided rich observations of what professors and students felt promoted critical-thinking skills, but did not note effective ways to construct a learning environment designed to encourage meaningful verbal exchanges among students. The quantitative data that the present study of test debate and test analysis produced, tested the efficacy of using cognitive dissonance facilitated by a conceptual debate providing a specific practice designed to produce the results Tsui sought.

Harvard professor Hunter Gehlbach’s (2007) research labeled this type of learning the acquisition of usable knowledge. Gehlbach asserted that, “While understanding the thoughts, feelings, and motivations of others—or social perspective taking—is crucial to successful relationships, researchers have found that it also plays an important role in the classroom” (2007, p. 1). Gehlbach explains, “One of the findings from my research is that the students who get higher grades also tend to be more motivated and more accurate in their perspective taking” (2007, p. 1). The opportunity to develop the skills necessary to

accurately see the perspectives of others closely relates to the constructs of educational theorists such as Dewey and Vygotsky. An environment designed to promote critical thinking requires student interaction in which all the members of the class, both students and teachers, are valued as thinkers. As students support their conceptual stance during test debate and test analysis, they are invested in the outcome, and therefore engaged in the learning of course content and concepts because it has a tangible value for them.

Neil C. Schmidt (2007) encapsulated the premise of critical thinking education by contending that effective classroom environments value the “attributes of thought, understanding, relating, judging, integrating, and reflecting [which] require deliberate, systematic and sustained attention” (p. 2). Schools that place thinking over the retention of facts seemed poised to develop students’ critical thinking skills. Both historical perspectives and recent research in the area of critical thinking education suggest that students need learning opportunities that require and value thinking if they are to develop the dispositions of good critical thinkers. The research on the effects of test debate and test analysis as tools to promote critical-thinking on the part of students, strives to identify and test the effects of a specific educational practice designed to bolster students’ critical thinking skills contributing to the existing research in the area of critical-thinking education. Historical constructs and recent research consistently laud the merits of developing thinking in students, yet few examples of empirically tested educational practices exist in the literature. This research study on test debate and test analysis filled some of the gap that exists between theory and practice in the area of critical-thinking education. The findings show that the practice of test debate and test analysis improved students’ critical thinking skills and literary analysis abilities by constructing a learning

environment that balances the acquisition of facts with the meaningful learning of course concepts.

Limitations of the Study

The most significant limitation of the study was the specific, detailed processes required for effective use of the test debate and test analysis processes. Research participants required extensive staff development for teachers and detailed practice for students. Narrative descriptions of the processes of test debate and test analysis often left research participants confused, and the elaborate training required a significant investment on the part of all members of the educational system.

Once teachers witnessed first-hand the powerful discourse generated during test debate, they embraced the process. An obvious limitation is access to classrooms in which test debate and test analysis are practiced. The clarifying nature of witnessing these processes in action is an essential part of widespread implementation, yet a practical way to institute teacher and student observation of test debates is a significant limitation. Certainly, the research findings presented in this study will serve as a foundation to entice teachers to investigate the process, but meaningful implementation will require a significant investment of educational resources.

A significant limitation is imposed by the need for test debate and test analysis trainers. Staff developers training teachers in the process must be experienced enough to assuage teachers' fears and guide teachers through the process. This undoubtedly requires a long-term commitment on the part of school districts. For example, during the research process, teachers in the treatment group struggled with the notion that control of the right and wrong answer would be guided by students' discussions. The idea of temporarily

forfeiting product for process is a legitimate concern only overcome by successful completion of the entire process.

Threats to External Validity

The external validity of this study is directly connected to teacher preparation prior to implementing test debate and test analysis. The generalizability of the findings suggests that similar results would be achieved in suburban middle and high schools, but the assertion that the process of test debate and test analysis will produce statistically significant improvement in other educational settings requires further research. Because the population was drawn from an experimentally accessible population, it is valid to generalize these research findings from the 157 participating students to the middle and high school students in the school from which the sample was taken. Yet, it might be presumptuous to generalize these findings to a dissimilar setting (Bracht & Glass, 1968).

Threats to Internal Validity

Further limitations are imposed by the extraneous variables that stem from the inherent differences in the way teachers conduct their classes. There was an attempt made to control extraneous variables, which threaten the internal validity of this study of test debate and test analysis, through the use of the three-group designs. The three-group designs sought to insulate this study from the effects of teacher disposition. Limitations also arise from factors such as student maturation and testing familiarity (Gall, Borg, & Gall, 1996). Students may have become more cognitively able during the duration of the treatment, which may have impacted the data set. The similarities between the pretests and the posttests also could have led to improvement emanating from the students' familiarity with the instruments rather than the effect of the treatment.

Implications

This study provided support for the use of test debate and test analysis as a way to improve students' critical-thinking skills and literary analysis abilities. The use of these processes had a statistically significant effect on students' literary analysis abilities measured using grade appropriate national and New York State standardized assessments, and on students' critical thinking skills measured using the *California Measure of Mental Motivations* (CM3). This section will discuss the degree to which the effect of the treatment was observed on two-group and three-group models using a multivariate analysis of variance with a covariate.

Implications of Effects of the Two-Group Design

Persons scoring high on Creative Problem Solving approach problem-solving situations with innovative or original ideas and solutions. These students “pride themselves on their creative nature, and this creativity is likely to manifest itself by a desire to engage in challenging activities” (Giancarlo, 2006, p. 5). The treatment group's participation in the test debate and test analysis processes led to a statistically significant scores on Creative Problem Solving due largely to the rigorous nature of the process, which requires students to seek creative ways to resolve the conceptual dilemmas posed in the literature studied in English classes.

Students engaged in the test debate and test analysis processes developed better problem-solving skills characterized by purposeful, self-regulatory judgment (Delphi Report, 1990). As the processes of test debate and test analysis unfold, students must find ways to formulate a stance on the author's purpose by incorporating their own perceptions with those of their classmates. The problem-solving aspect of the process

requires students to substantiate assertions about literary text through the manipulations of specific events and literary elements employed by authors. The processes develop students' ability to synthesize information from a variety of sources: their own initial interpretation, class lectures and discussions, and the asserted contradictory stances of their classmates.

The Creative Problem Solving scale of the CM3 asserts that students scoring high on this scale embrace activities that require innovation and challenge seeking behavior, inherent qualities of test debate and test analysis. The Innovation subscale of Creative Problem Solving identifies students who have a sense of confidence in their ability to solve difficult problems, and they tend to “identify alternatives and take creative and innovative paths to solve problems” (Giancarlo, 2006, p. 5). Students in the treatment group participated in a process that required them to activate existing knowledge while forming a final stance on the conceptual dilemmas posed by the process. The treatment process develops the quality of innovation as students seek ways to solve the problems posed by this assessment technique, problems which transcend the very nature of the course. Students are asked to think critically while suspending final judgment until the final debate ferrets out several possible paths towards resolving complicated questions stemming from the literature.

The challenge-seeking component of Creative Problem Solving is defined as a preference for challenging and complicated activities (Giancarlo, 2006). Creative Problem Solving scores suggest that students scoring high on this scale seek to increase their knowledge base and skills, an educational goal sought by most educational systems. The purpose of education is not to merely transfer a body of existing knowledge that

already exists, but to develop students' ability to think and resolve the difficult problems that face society. The test debate and test analysis processes are a viable method for developing students as thinkers with the skills needed to address the challenges imposed by an ever-changing global community.

In the tests of Between-Subjects Effects, the partial Eta-squared effect size for Creative Problem Solving was a modest 4.2%, still impressive considering the short duration of the research study. Dispositions, such as critical thinking, develop slowly over time. As Orr and Klein (1991) explained, the inclinations of critical thinking are "dimensions of one's personality which relate to how likely a person is to approach problem identification and problem solving by using reasoning" (p. 133). Even small changes in dimensions of personalities suggest that test debate and test analysis had an important impact on the treatment group. A more comprehensive implementation of the process of test debate and test analysis in all academic disciplines over students' entire educational careers would likely produce more substantial effects. As Giancarlo explained, "the dispositional domains measured by Creative Problem Solving are not linked with any particular curricular area" (Giancarlo, 2006, p. 4). The process of test debate and test analysis have the potential to provide teachers with a means to develop critical thinking skills that students can activate in any academic discipline.

Students in the treatment group also had statistically significant higher scores on Scholarly Rigor than the no treatment group, a scale of the CM3 defined as the "disposition to work hard to interpret and achieve a deeper understanding of complex or abstract material" (Giancarlo, 2006, p. 10). The scores of the treatment group differed from the no-treatment group. The treatment group's participation in test debate and test

analysis promoted the disposition to comprehensively seek new knowledge and examine new content in depth (Giancarlo, 2006). The literature on critical thinking suggests that there is a connection between critical thinking dispositions in the prediction of academic success (Bachman, 1999). Test debate and test analysis promotes critical thinking, a disposition necessary for successful educational growth.

In the tests of between-subjects effects, the partial Eta-squared effect size for Scholarly Rigor was 4.0%, which was consistent with the effect size of Creative Problem Solving. Again, a multidisciplinary, long-term implementation of the process of test debate and test analysis would likely increase the effect size, but considering the limited duration of the study, this dispositional change suggests that the process has great promise for the development of students' critical thinking skills.

The most significant impact of the process of test debate and test analysis was observed in the treatment group's Literary Analysis scores. The statistical significance reveals that the process of test debate and test analysis greatly improved students' ability to answer literature comprehension questions that make up grade specific national and New York State standardized tests. Test debate and test analysis develops dispositions that impact students' scores on assessments that are often used to measure the success or failure of school systems. As students used test debate and test analysis assessments, they developed a metacognitive approach to answering test questions, which positively impacted future test performance.

The effect size for Literary Analysis variable was an impressive 70.2%. These results can be attributed to the direct relationship between the process of test debate and test analysis and the test taking skills necessary for successful completion of the

standardized literature comprehension tests. It is important to note that the treatment group was comprised of a sample made up of heterogeneously grouped sixth grade students and homogeneously grouped eleventh grade Advanced Placement English Language students. It can be concluded that the process of test debate and test analysis produced the desired results across a broad developmental range. The transferability of the process suggests that implementation is warranted as noted in the Future Research section of this dissertation.

Implications of the Effects of Three-Group Design

The three-group multivariate analysis of variance with a covariate sought to test the impact of teacher training on the final data set. These data supported the contentions that the treatment itself led to the differences in the groups rather than the training of the teachers. Teacher training did not impact the students' performance. Across all six dependent variables, students' mean scores in the teacher not trained no treatment group equaled or exceeded the mean scores of the trained no treatment group. There were no statistically significant differences in mean scores between the groups taught by teachers trained in test debate and test analysis and the groups taught by teachers not trained in the processes. Teacher training appeared to have no additional impact effect on the students' performance.

The elimination of this important extraneous variable (teacher training) further supports the widespread implementation of the process of test debate and test analysis. The data shows that the growth in students' critical thinking and literary analysis ability can be attributed to the efficacy of the process rather than training of the teacher. The processes of test debate and test analysis led to observable student growth.

Suggestions for Future Research

There are extensive theoretical constructs supporting the need for an educational focus on critical thinking. Dewey, Vygotsky, Piaget, and Ausubel built their theories around a central premise that meaningful learning is constructed through a process of metacognitive reflection. The veracity of these constructs is generally accepted, yet meaningful research testing specific educational practices that foster critical thinking is scarce. Tsui (2002) contended that, “As yet, however, little substantiated knowledge on effective pedagogy comes from research on critical thinking. Very few studies on critical thinking ... examine the impact of instructional factors” (p. 741). Tsui’s assertions were supported by J. H. McMillan’s (1987) conclusion that “Little consistency emerges from the empirical research literature as to specific instructional techniques that effectively enhance students’ ability to think critically” (p. 3).

There is little debate over the need for critical-thinking education. Future research must seek empirical evidence supporting the use of specific educational practices appraising classroom activities through qualitative and quantitative means. This section will recommend four areas for future research in the area of critical thinking education: (a) qualitative studies focusing on students’ attitudes, (b) qualitative studies focusing on teacher attitudes, (c) longitudinal studies on critical thinking, and (d) studies focusing on specific classroom activities.

Qualitative Research and Students’ Attitudes

Few studies exist that use qualitative measures to assess the impact of classroom activities on critical thinking skills (Tsui, 2002). Future research in the area of critical thinking should use qualitative measures to investigate the effects of critical thinking

pedagogy on students' metacognitive processes. As students participate in classroom activities, the conceptual connection between the activity and their thinking must be monitored to determine if the prescribed treatment is truly impacting students' thinking. Everett and Zinser's (1998) research found that some students resented expending the cognitive energy required to participate in activities that required active participation and ongoing metacognitive reflection. Qualitative studies of specific teaching methods that focus on students' attitudes would provide invaluable information in the area of critical-thinking education. Through the use of surveys and semistructured interviews, future researchers may discover insights that guide traditional educational practices toward reform designed to promote students' critical thinking development.

Qualitative Research and Teachers' Attitudes

Future research must also address the area of critical thinking as it affects teachers' attitudes. All meaningful educational reform starts with staff development, so researchers must monitor the impact of teacher training on teachers' attitudes and behavior. The information gained by conducting a qualitative appraisal of teachers' attitudes as they attempt to implement new teaching strategies will provide important information on the challenges and frustrations that teachers face. Research studies using interviews and classroom observations might clarify some of the issues that have hindered the progress of critical thinking education. The information gleaned from a study of teachers' attitudes is necessary to see how educational systems can balance the pressures to teach students to be better thinkers with the increasing pressures applied by the influx of standardized assessments.

Longitudinal Studies on Critical Thinking

Critical thinking is a disposition that develops over time. Research that monitors the improvement over time of critical thinking skills in students would greatly enhance the literature on critical thinking. A periodic analysis of students' critical-thinking skills that produced longitudinal data would provide educational systems with the empirical evidence they need to set attainable long-term goals in the development of students' critical thinking skills. The snapshot view of this study executed over a four-month treatment period provides insights and suggests in the area of critical thinking; a more extended longitudinal study designed to evaluate the progress of students' critical thinking skills would greatly augment the existing literature.

Studies Focusing on Specific Classroom Activities

There is no shortage of theories in the areas related to critical thinking, and these theories provide the impetus for meaningful educational reform. It is, however, important to build a body of research that tests the effectiveness of specific educational strategies in the area of critical thinking. Future research should seek evidence that certain classroom techniques support the acquisition of critical thinking skills in students in a variety of educational settings across many academic disciplines. As the Boyer Commission (1998) contended with regard to undergraduate education, teachers must seek ways to “imbue students with a sense of excitement of discovery and opportunities for intellectual growth” (p. 20). These educational practices will serve as models worth emulating if they are supported by empirical data suggesting they are effective. This study of test debate and test analysis provides quantitative data that this educational practice enhances

students' ability to think and reason beyond what is commonly offered by the traditional educational model.

Summary

Test debate and test analysis is a specific classroom activity that promotes critical thinking skills and improved literature comprehension on the part of students. There is a substantial body of literature in the area of critical-thinking education that supports the assertions made by this study, but a review of recent studies reveals a need for empirical research that tests the effectiveness of specific classroom activities designed to improve students' ability to metacognitively think and reason. This study linked the theoretical underpinnings of critical thinking education with a practical process that can be replicated. Despite the need for future research, this study supports the use of test debate and test analysis as a viable way to enhance students' critical thinking and literary analysis abilities.

As teachers search the existing body of theoretical literature, they are rewarded with sound constructs that suggest the need for educational reform that values critical thinking over mere memorization of facts. Educators pursuing critical-thinking models soon realize what is missing from the discussion: specific educational practices that promote critical thinking skills and dispositions. This study on the impacts of test debate and test analysis on students' critical thinking and literary analysis ability provides a partial solution and supports the implementation of these processes of test debate and test analysis as part of educational reform striving to develop students as thinkers.

References

- Ausubel, D. P. (1962). A subsumption theory of meaningful verbal learning. *Journal of General Psychology*, 66, 213-224.
- Bachman, M. L. (1999). Anxiety, critical thinking and age as performance predictors of community college nursing students. *Dissertation Abstracts International*, 60(3-B): 1026.
- Boyer Commission on Educating Undergraduates in Research University. (1998). *Reinventing undergraduate education: A blueprint for America's research universities*. Stony Brook, NY: State University of New York.
- Bracht, G. H., & Glass, G. V. (1968). The external validity of experiments. *American Educational Research Journal*, 5, 437-474.
- Bray, J. H., & Maxwell, S. E. (1985). *Multivariate analysis of variance*. Beverly Hills, CA: Sage.
- Brooks, J. G., & Brooks, M. G. (1993). *In search of understanding: The case for constructivist classrooms*. Alexandria, VA: Association For Supervision and Curriculum.
- Cole, M., & Scribner, S. (1978). Introduction. In L. S. Vygotsky, *Mind in society*. Cambridge, MA: Harvard University Press.
- The College Board. (2007). *The AP exam*. Retrieved August 21, 2007 from http://64.233.169.104/search?q=cache:59_y83_4t-OJ:apcentral.collegeboard.com/apc/public/exam/index.html+%22To+ense+that+AP+Exams+accurately+measure+college-

- level+knowledge+and+performance+in+each+discipline&hl=en&ct=clnk&cd=1
&gl=us
- Davydov, V. V. (1986). *The problem of development-generating learning*. Moscow: Pedagogika.
- Delphi Report. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. ERIC Doc. No. ED 315-423.
- DeMauro, G.E. (2001). *Construct properties of the New York State English language examinations*. (Albany, NY). Office of State Assessments.
- Dewey, J. (1938). *Experience and education*. New York: Simon & Schuster.
- Donald, J. G. (1983). Knowledge structures: Methods for exploring course content. *Journal of Higher Education*, 54, 31-41.
- Driscoll, M. (2000). *Psychology of learning for instruction*. Boston: Allyn & Bacon.
- Everett, M. D., & Zinser, O. (1998). Interdisciplinary social science courses: Using a critical thinking approach. *Journal of General Education*, 47, 225-241.
- Facione, P. A. (1991). *Using the California Critical Thinking Skills test in research, evaluation and assessment*. Millbrae, CA: Insight Assessment.
- Friedman, T. L. (2005). *The world is flat*. New York: Farrar, Straus and Giroux.
- Furedy, C., & Furedy, J. (1985). Critical thinking: Toward research and dialogue. In J. Donald & A. Sullivan (Eds.), *New Directions for Teaching and Learning*, No. 23 (pp. 51-69). San Francisco: Jossey-Bass.
- Gall, M. D., Borg, W. R. & Gall J. P. (1996). *Educational research: An introductory*, (6th ed).. New York: Longman.

- Gehlbach, H. (2007). *Usable knowledge. Social perspective taking: A multidimensional approach*. Cambridge, MA. Harvard Graduate School of Education. Retrieved August 17, 2007 from: www.uknow.gse.harvard.edu/teaching.
- Giancarlo, C. A. (2006). *The California Measure of Mental Motivation: User manual*. Millbrae, CA: California Academic Press.
- Giancarlo, C. A., Blohm, S. W, & Urdan, T. (2004). Assessing secondary students' disposition toward critical thinking: Development of the California Measure of Mental Motivation. *Educational and Psychological Measurement*, 64, 347-364.
- Giancarlo, C. A., & Facione, P. A. (2000). *The California Measure of Mental Motivation*. Millbrae, CA: California Academic Press.
- Green, S. B., & Salkind, N. J. (2005). *Using SPSS for the Windows and Macintosh: Analyzing and understanding data* (4th ed.). Upper Saddle River, NJ: Prentice Hall.
- Gunn, T. M., Grigg, L. M. & Pomahac, G. (2006). *Critical thinking with gifted and talented adolescents*. University of Lethbridge, AB. Paper presented at WestCast, 2006.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis*. Englewood Cliffs, NJ: Prentice Hall.
- Harris, D. M. (1998). *Effects of metacognitive skill training upon academic performance in accounting*. Unpublished doctoral dissertation, The University of Texas at Austin, Austin, Texas.

- Karpov, Y. V. & Bransford, J. D. (1995). L. S. Vygotsky and the doctrine of empirical and theoretical learning. *Educational Psychologist*, 30(2), 61-66.
- Lee, M. C. (2003). *Impacts of cognitive structuring methods on students' critical thinking enhancement in on-line collaboration learning*. Unpublished doctoral dissertation, University of Illinois at Chicago, Chicago, Illinois.
- Leopold, W. (1997, March 4). Educators target teaching "Johnny" to think better. Turning students into competent analysts is the goal. *Chicago Tribune*, p. A6.
- Lynd-Balta, E. (2006). Using literature and innovation assessments to ignite interest and cultivate critical thinking skills in an undergraduate neuroscience course. *CBE Life Science Education*, 5, 167-174.
- McMillian, J.H. (1987). Enhancing college students' critical thinking: A review of studies. *Research in Higher Education*, 26, 3-29.
- Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied multivariate research: Design and interpretation*. Thousand Oaks, CA: Sage.
- Nordvall, R. C., & Braxton, J. M. (1996). An alternative definition of quality of undergraduate college education. *Journal of Higher Education*, 67, 483-497.
- Norris, S. P. (1985). Synthesis of research on critical thinking. *Educational Leadership*, 42(8), 40-45.
- Orr, J. B & Klein, M. F. (1991). Instruction in critical thinking as a form of character education. *Journal of Curriculum and Supervision*, 6, 130-144.
- Pantina, N. S. (1957). Formation of writing skills based on the type of task orientation. *Voprosy Psikhologii*, 4, 117-132.

- Piaget, J. (1969). *Science of education and the psychology of the child*. New York: Viking.
- Quitadamo, I. J., & Kurtz, M. J. (2006). Learning to improve: Using writing to increase critical thinking performance in general education biology. *CBE Life Sciences Education* 6, 140-154.
- Reynolds, G. F. (1946). Oral interpretation as graduate work in English. *College English*, 11, 204-210.
- Schmidt, N. C. (2007). *What are schools for?* Retrieved January 12, 2007 from www.leapfrog.com.
- Schwarm, S., & VanDeGrift, T. (2002). *Using classroom assessment to detect students' misunderstanding and promote metacognitive thinking*. Retrieved July 7, 2007 from: <http://www.cs.washington.edu/research/edtech/publications/sv02-icls.pdf>.
- Stevens, J.P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Hillsdale, NJ: Erlbaum.
- Talyzina, N. F. (1981). *The psychology of learning*. Moscow: Progress.
- Thin, T. O. (2001). *A qualitative inquiry into the critical thinking process of hospitality management students*. Unpublished doctoral dissertation, Iowa State University, Ames, Iowa.
- Tishman, S., & Andrade, A. (1995). *Thinking dispositions: a review of current theories, practices, and issues*. ACCTION report #1. Washington, DC. Available from <http://learnweb.harvard.edu/alps/thinking/docs/Dispositions.htm>
- Tsui, L. (2002). Fostering critical thinking through effective pedagogy: Evidence from four institutional case studies. *The Journal of Higher Education*, 73, 740-763.

- Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press.
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Walzer, A. E. (1991). The meaning of “purpose.” *Rhetoric Review*, 10(1), 118-129.
- Wellman, R. R. (1970). Socratic method in xenophon. *Journal of the History of Ideas*, 37, 307-318.
- Zuber-Skerrit, O., & Fletcher, M. (2007). The quality of an action research thesis in the social sciences. *Quality Assurance in Education*, 15, 413-436.

Appendix A:

Student and Parent/Guardian Consent Form

Appendix A:

Student and Parent/Guardian Consent Form

Dear Student and Parent/Guardian:

As part of my doctoral research at Western Connecticut State University, I will be studying the effects of test analysis and test debate on students' critical thinking skills, literary analysis, attitudes, and behaviour. I will be using student surveys, interviews, a critical thinking questionnaire, and National and State assessments to determine the effectiveness of the educational practice of test analysis and test debate.

If students and parents agree to participate, students will be asked to complete *The California Measure of Mental Motivation (CM3)*, student surveys, interviews, and National and State assessments as a way to evaluate test analysis and test debate. Students may refuse to answer any question on either the survey or interview, and they are free to withdraw at any time. If students or parents/guardians do not wish to participate in this study, this will not have any effect on students' grades or any aspect of their academic record.

To protect students' privacy, the students' will be assigned a number, which will be used to match data. Teachers will not have access to students' replies, and all data will be kept in a secure location; all published results of this research will not contain any students' names making it impossible to identify any student.

Students and parents/guardians should sign the attached consent form if they are willing to participate. Please return the signed consent forms directly to the students' English teacher. If students or parents/guardians have any further questions about the study, please do not hesitate to contact me.

Sincerely,

To be completed by Student

I have read and understand the conditions, and I consent to voluntarily participate in this research study. I realize that I am free to withdraw my consent at any time without any negative consequences.

Student's Name (Please Print):

Signature: _____

To be completed by Parent/Guardian

I have read and understand the conditions, and I consent to the voluntarily participation of my child in this research study. I realize that I am free to withdraw my consent at any time without any negative consequences.

Parent/Guardian Name (Please Print):

Signature: _____

Relationship to Student: _____

Phone number: _____ Email address: _____

Appendix B:
District Permission Form

Appendix B:

District Permission Form

December 5, 2006

WCSU Institutional Review Board:

The district grants Nicholas Kowgios permission to conduct the study: “Effects of Conceptual Assessments Using Test Analysis and Test Debate on Critical Thinking Skills, Attitudes, and Literary Analysis” as part of his doctoral research for Western Connecticut State University. The District recognizes that the study will use a sample ($N=157$) of North Salem students, and that no students’ names will be part of the study.

Appendix C:
Human Subjects Research Review Form

Appendix C:

Human Subjects Research Review Form

WESTERN CONNECTICUT STATE UNIVERSITY

Human Subjects Research Review Form

Principal Investigator: Nicholas Kowgios

New research project Continuation ____ Modification ____ Teaching ____

____ Exempt Review (attach a completed copy of the “Application for Exemption”)

Expedited/Full Review

To complete this form, please follow the instructions in sections A and B.

Checklist for attachments:

- ____ Completed Application for Exemption (if claiming exemption)
- ____ Answers to A1 through A 6
- ____ Survey or questionnaire
- ____ Informed consent form
- ____ Student’s current NIH training certificate
- ____ Instructor’s current NIH training certificate
- ____ Chair’s current NIH training certificate

The **department chair** and the **principal investigator (PI)** must sign this form. If the PI is a student, his/her faculty supervisor must also sign.

Assurance of continued compliance with regulations regarding the use of human subjects. I certify that the information provided for this project is accurate. If procedures for obtaining consent of subjects change, or if the risk of physical, psychological, or social injury increases, or if there should arise unanticipated problems involving risk to subjects or others, I shall promptly report such changes to the Institutional Review Board. I shall report promptly unanticipated injury of a subject to my department chair and to the Institutional Review Board.

Principal Investigator’s Signature

Date

Faculty Supervisor’s Signature (if PI is a student)

Date

Department Chair’s signature

Date

=====

Committee Action:

_____ Approved through exempt review
committee review

_____ Approved by full

_____ Approved through expedited review

_____ Not approved;
clarification or
modification required

IRB Chair's Signature

Date

A. Instructions for completing the HUM-1 Form (attach answers):

1. Describe the characteristics of the subject population (anticipated number, age ranges, gender, ethnic background, and health status).
The target population will be a group of students in grades 7 through 11 where N = 267. Research will be conducted at a small, suburban school district with a total student population of 780 students. Demographically, the student population's socio-economic backgrounds are middle to upper class with a median home income of \$174,000.00. The gender breakdown is 54% female and 46% male. There is a small percentage of ethnic diversity that includes 96% white students, 2% Hispanic students, 1% black students, and 1% of the students listed in the *other* category. The target sample is representative of the school population in gender makeup and ethnicity.
2. Explain the rationale for use of special classes of subjects (children, mentally disabled, elderly, prisoners, or others).
N/A
3. Identify the records or data to be obtained for individually identifiable living human subjects. Students' level of critical thinking skills will be assessed using the California Measurement of Mental Motivation (CM3) (Giancarlo & Facione, 2000). The internal consistency for the 25-item CM3, evaluated using Cronbach's alpha coefficient, was .53 to .83 for the 4 scales: Learning Orientation, Creative Problem Solving, Mental Focus, and Cognitive Integrity (Giancarlo & Facione, 2000). Student surveys and questionnaire schedules are included in the study. This study will use grade specific New York State English Language Arts assessments and the English Language and Composition Advanced Placement part 1 raw scores to measure literary analysis, and the CM3 to measure critical thinking.
4. Describe plans for recruitment of subjects and the consent procedures to be followed, or explain why consent is not needed.

The participants in this study will be a *sample of convenience* selected to suit the purpose of the study. The students in the sample classes will be given an informed consent form to be completed by their parents.

- 5. Describe safeguards to assure anonymity and voluntary participation of subjects. In the case of student subjects, indicate that failure to participate in or withdrawal from the project will not affect class grade.

Names will not appear on any documents, and all participants in the study will be assigned a number for identification purposes. All documents will be stored in a secure location only accessible to the primary researcher. Failure to participate or withdrawal from this study will have no impact on students' grades.

- 6. "Subject at risk" means any individual who may be exposed to the possibility of injury, including physical, psychological, or social injury, as a consequence of participation as a subject in any research, development, or related activity that departs from the application of those established and accepted methods. [45CFR 46.3(b)]

N/A

B. Answer the following (if you answer yes to either question, the protocol requires full review):

- Does your project involve **risk of physical injury** to subjects?
 Yes No

(If yes, describe the nature of the risk, the justification for undertaking the risk, and the procedures used to obtain the subject's informed consent to take the risk.)

- Does your project involve **risk of psychological or social injury** to human subjects?
 Yes No

(If yes, describe the nature of the risk, the justification for undertaking the risk, and the procedures used to obtain the subject's informed consent to take the risk.)

NOTE: If participation in the research involves physical, psychological, and/or social risk to the subject, the informed consent form must say so in bold type.

=====

Please send the completed form (if the protocol requires full review, send 12 copies) to: Director of Grant Programs, 321 Warner Hall. If you have questions, call 7-8281.

Last updated 6/26/06