EVALUATING THE EFFECTS OF CREDIT-BASED TRANSITIONAL PROGRAMS ON HIGH SCHOOL STUDENTS' CRITICAL THINKING SKILLS

by

Jane M. Walther

A dissertation

submitted in partial fulfillment

of the requirements for the degree of

Doctor of Education in Curriculum and Instruction

Boise State University

Summer 2009

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the dissertation submitted by

Jane MacDonald Walther

Dissertation Title: Evaluating the Effects of Credit-based Transitional Programs on High School Students' Critical Thinking Skills

Date of Final Oral Examination: 10 June 2009

The following individuals read and discussed the dissertation submitted by student Jane Walther, and they also evaluated her presentation and response to questions during the final oral examination. They found that the student passed the final oral examination, and that the dissertation was satisfactory for a doctoral degree and ready for any final modifications that they explicitly required.

Keith Thiede, Ph.D.	Chair, Supervisory Committee
Kathleen Budge, Ed.D.	Member, Supervisory Committee
Anne Gregory, Ph.D.	Member, Supervisory Committee
Jennifer Snow-Gerono, Ph.D.	Member, Supervisory Committee

The final reading approval of the dissertation was granted by Keith Thiede, Ph.D., Chair of the Supervisory Committee. The dissertation was approved for the Graduate College by John R. Pelton, Ph.D., Dean of the Graduate College.

DEDICATION

I dedicate this work to my parents, Audrey and Charlie Walther, who instilled in me from the beginning a profound love of learning, and whose ever-enduring love and support has gotten me to where I am today; and to the late Genevieve and Dermot O'Flynn, and Irene and Augustus Crenshaw, without whom none of this would have been possible.

ACKNOWLEDGEMENTS

I wish to express my gratitude to my committee members, in particular to Dr. Keith Thiede, Committee Chair, for his insight, patience, and sense of humor throughout this process. Many thanks are also extended to Dr. Kathleen Budge, Dr. Anne Gregory, and Dr. Jennifer Snow-Gerono, each of whom provided advice and encouragement.

In addition, I would like to thank JoJo Brunelle, Bob Carignan, and Derek McCormick for their invaluable assistance in getting this study off the ground; the Bieter family, whose generous donation helped finance the pursuit of this degree; and my friends from the doctoral program, especially Julie Dahlgren and Diana Hooley, who spent countless hours listening, discussing, and sharing insight that added to the depth of this study.

Special thanks go out to Drs. Lois and Roland Worthington, who encouraged me to enroll in the doctoral program and provided support throughout the four years of study; to Dr. Christine Isaacs, for her confidence and understanding; to Jim Harper, for his friendship and encouragement; and to Ness Kraikhaijornkiti, whose proverb inspired me every day.

Finally, thank you to my family: to my parents, Audrey and Charlie Walther, and to my siblings and siblings-in-law: Tom and Katie, Nancy and Pat, and Susan and Jamie, for their support throughout this tremendous undertaking; and to the little ones: Charlie, Ryan, Bridget, Eamon, Megan, and Stella, who always give me something to smile about.

ABSTRACT

This study compared the critical thinking (CT) skills of Grade 12 students (N = 60) enrolled in different academic programs: International Baccalaureate (IB), Advanced Placement (AP), and a control group (NON). CT was assessed using the Ennis-Weir Critical Thinking Essay Test, which provided measures of five CT subscales: Evaluation of Argument, Deduction, Inference, Recognition of Argument, and Interpretation; and a separate category of CT skill: Use of Emotive Language to Persuade.

Despite similar demographics across groups, there were significant differences in three of the five subscales of CT and Use of Emotive Language to Persuade. IB and AP scores were higher than NON scores in Inference, Recognition of Assumption, and Interpretation. IB and AP scores did not differ across the five subscales; however, IB scores were higher than AP and NON scores in Use of Emotive Language to Persuade. AP and NON scores did not differ in this category.

These findings suggest that academic programming may have an effect on the development of students' CT skills. However, this study did not examine instruction within the programs; therefore, it is not possible to attribute the reported differences solely to the programs.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

Overview

This introductory chapter discusses the importance of CT skills in education and the need for more effective teaching of these skills. It presents research questions indicating intentions to test the influence of academic programming on students' CT skills and describes the relevance and use of the study's instrument. This chapter provides definitions of important terms used in the study. It outlines the study's delimitations and limitations, and the measures taken to address them.

Statement of the Problem

Developing critical thinking (CT) skills is one of the most important aims of education (Bernard, Zhang, Abrami, Sicoly, Borokhovski & Surkes, 2008; Ennis, 1993; Erwin & Sebrell, 2003; Giancarlo, Blohm & Urdan, 2004; Noddings, 2006, 2008; Siegel, 1986; ten Dam & Volman, 2004; van Gelder, 2005; Willingham, 2007). Acquisition of these skills is vital because they equip us with the capacity to participate in a democratic society (ten Dam & Volman, 2004), make complex choices (Dolhenty, 2008; Lampert, 2006), commit to social justice (Bickford, 2008), and reflect on our beliefs and actions (Jackson, Ignatavicius, & Case, 2006). Every student has the right to an education that develops these important skills (Siegel, 1986).

Despite the importance of teaching CT, studies have shown that most schools fall short in challenging students to think critically and develop the reasoning skills necessary to grapple with the complexities of modern life (Paul, 1993; Reed, 1998). This may be due in part to lack of a clear idea as to what constitutes CT (Paul, 1993); instruction that focuses more on the memorization of facts than engagement in deep thinking (Gallagher, 1998); and/or high-stakes assessment that undermines CT (Noddings, 2004).

Purpose of the Study

This study attempted to contribute to the knowledge of CT by comparing the CT skills of students in different academic programs. Its purpose was to assess quantitatively the CT skills of high school students in the International Baccalaureate (IB), Advanced Placement (AP), or a more traditional (control) program, and to discern how these programs might affect students' CT development.

Although participants shared similar demographics across groups, it was not certain whether significant differences in scores could be attributed to academic programming, or if factors such as differing schools, teachers, and forms of instruction played a role.

Research Questions

This study sought to address the following questions:

1. What differences in critical thinking skills, if any, exist between students enrolled in academic programs that purport to foster critical thinking, namely the IB and AP programs, and students in a control group?

2. What differences in critical thinking skills, if any, exist between students enrolled in the IB or AP program, both of which purport to foster critical thinking?

Definitions

The following terms are defined for use in this study.

Advanced Placement (AP): A credit-based transitional program designed for Grade 12 students. Individual AP courses may be taken in one or more subject areas in which a students excels.

Credit-based Transitional Program: An academic program designed to offer college-level curriculum to high school students so that they may receive accelerated instruction, earn college credit prior to entering college, and develop study skills necessary for college (Bailey & Karp, 2003). Two well-known examples are the AP and IB programs (Bailey & Karp, 2003).

International Baccalaureate (IB): A credit-based transitional program taken during Grades 11 and 12. IB courses are taken in native language, second language, individuals and societies, experimental sciences, math and computer science, and the arts. The program includes three core components: Theory of Knowledge (a philosophy course); Extended Essay (a research project); and CAS (150 hours of Creativity, Action, and Service).

Socratic Questioning: An instructional technique used to encourage deep thinking and discourse by asking questions so that students may find the desired knowledge in their answers.

Delimitations and Limitations

Sixty Grade 12 students from two different schools in a mid-sized city in the northwestern United States constituted the research sample. Because only 20 IB students participated, 20 matched participants were selected for the AP and control group. This sample size is relatively small and one test was administered once; therefore, results will not generalize to a larger population of high school students that does not share similar demographic characteristics.

One limitation of this study was the short time during which it took place. One measurement during one 50-minute period provides an indication rather than an in-depth look at students' CT skills.

Another limitation resides in the fact that there were no non-IB students to recruit at Site A, as all students were enrolled in the IB program. Consequently, this study was limited to having only one control group, which consisted of non-AP students from Site B. Inclusion of non-IB students would have allowed the researcher more room to speculate if perhaps differences in schools (i.e., teachers and instruction), more so than in programs, affected students' scores.

This study includes information indicating that the IB and AP programs strive to develop CT skills (The AP Advantage, n.d.; Barbour & Streb, 2008; International Baccalaureate Organization (IBO), 2008); however, it was not possible to address if/how these particular participants—be they IB, AP, or students in neither program—were (or were not) taught to think critically.

Another limitation is found in the differences between the two schools included in the study. Site A was a private, independent school, while Site B was a private, Catholic

school. It may be possible that in light of these differences, instruction was different based on the visions of the two schools, with the former focusing on academic excellence, innovative thinking, and creative self-expression, the latter on academic excellence, moral leadership, and social responsibility in allegiance to the Catholic church.

Another limitation includes the difference in duration of the IB and AP programs. The IB is a two-year program; AP courses are designed for a period of one year. The longer period of instructional time may favor IB students; however, it is likely that AP students had received prior instruction focused on preparing them for AP coursework.

Another potential limitation was the researcher's position as a former IB teacher at Site A. Although measures were taken to ensure objectivity, (e.g., scoring exams blindly), it is possible that her past teaching experience, knowledge of the IB program, and familiarity with students at Site A may have affected her role in conducting the experiment.

The remaining limitations of this study were found in the instrument. Although the EW was designated the most appropriate commercial test of CT for this study, its focus (overnight street parking) may have been irrelevant to some participants; consequently, these students may not have put forth the effort required to earn a high score on the test. In addition, due to a lack of normative data on the EW, the results of this study are limited to how different groups of students faired relative to each other, rather than in comparison to other, similar populations. Finally, the EW testing manual does not provide examples of students' work to aid in scoring tests. All tests were scored blindly with an inter-rater reliability of .86, but additional sources indicating what the authors of the test considered excellent, fair, and poor reasoning, would have further

clarified the scoring process, thus yielding results more consistent with the authors' objectives.

CHAPTER TWO: LITERATURE REVIEW

Overview

Chapter Two presents research on definitions, instruction, and assessment of critical thinking (CT). It begins by examining definitions of CT, including exploration of CT as a disposition. It explores effective ways of teaching CT. It discusses multi-aspect assessment of CT and compares a multiple-choice test, *The Watson-Glaser Critical Thinking Appraisal,* with an essay test, *The Ennis-Weir Critical Thinking Essay Test.* The benefits of essay tests and the alignment of essay tests with the principles of CT are explored.

Critical Thinking Definitions

Due to the complexities inherent in thinking, a collective definition of CT does not exist (Paul, 1993); however, different definitions of the term share common characteristics. This review of literature explores CT in terms of "reflective thinking," "creative thinking," and "higher-order thinking" and discusses how these concepts are used in place of or in conjunction with definitions of CT.

Reflective Thinking

Dewey's (1909) defined CT as "the active, persistent, and careful consideration of a belief or supposed form of knowledge in the light of the grounds which support it and

further conclusions to which it tends" (p. 6). Use of the term "consideration" in this definition suggests a reflective element in CT.

Ennis provides a similar definition, with a more overt reference to reflective thinking. According to Ennis (1989), CT is "reasonable, reflective thinking that is focused on deciding what to believe or do" (p. 4).

Paul distinguishes his definition from Ennis's by incorporating more explicitly the concept of metacognition, the idea of "thinking about one's thinking,...and consciously aiming to improve it by reference to some model of good thinking in that domain" (Paul, Fisher & Notsch, 1993, p. 4). He states:

> Critical thinking is that mode of thinking—about any subject, content or problem—in which the thinker improves the quality of his or her thinking by skillfully taking charge of the structures inherent in thinking and imposing intellectual standards upon them. (p. 4)

Though less explicit, Jackson, Ignatavicius and Case's (2006) definition discusses metacognition, as it suggests that critical and reflective thinking are reciprocal. They explain:

> As we develop the ability to think critically we also develop the ability to participate in self-reflection. It is only by critical thinking and reflecting on what we think and believe and how we behave in the world that we can come to know what our attitudes about life truly are. Learning to think critically about our attitudes allows us to identify them and then decide if we want to change them or keep them. (p. 26)

Noddings (2006) emphasizes that she rarely differentiates CT from reflective thinking. She asserts that CT is personal; it "refers not only to the assessment of arguments...but also to the diligent and skillful use of reason on matters of moral/social importance—on personal decision making, conduct and belief" (p. 4). Noddings underscores the idea that CT goes beyond assessing arguments logically. Noddings

suggests that it is a conscious act of carefully reflecting on something, combined with reason, which results in choices as to how to act and what to believe.

Creative Thinking

An anonymous quote describes the relationship between creative and CT: "The critical and creative functions of the mind are so interwoven that neither can be separated from the other without an essential loss to both" (as cited in Paul & Elder, 2004, p. 3). Likewise, it has been asserted that "critical and creative thinking are interrelated and complementary aspects of thinking. Almost all of the thinking which we undertake contains some critical and some creative aspects" (Saskatchewan Education, n.d.). This connection between creative and CT is underscored by Paul (1993):

> "Criticality" and "creativity" have an intimate relationship to the ability to figure things out. There is a natural marriage between them. Indeed, all thinking that is properly called "excellent" combines these two dimensions in an intimate way. Whenever our thinking excels, it excels because we succeed in designing or engendering, fashioning or originating, creating or producing results and outcomes appropriate to our ends in thinking. It has, in a word, a *creative* dimension. (p. 102)

Jackson et al. (2006) contend that a large part of CT involves creativity; they suggest that CT "becomes part of the (creative) process, as we seek formalization and a clearer understanding of the applications of the creative idea" (p. 8).

> Paul and Elder (2004) describe a strong connection between creativity and CT: The very definition of the word "creative" implies a critical component (e.g. "having or showing imagination and artistic intellectual inventiveness"). When engaged in high-quality thought, the mind must simultaneously produce and assess, both generate and judge the products it fabricates. In short, sound thinking requires both imagination and intellectual standards. (p. 4)

Creative thinking and CT are interrelated, as illustrated by the previous explanations; however, it is held that creative thinking differs from CT in that it is "involved with the creation or generation of ideas, processes, experiences or objects; [whereas] critical thinking is concerned with their evaluation" (Saskatchewan Education, n.d.). This differentiation is echoed by Paul and Elder (2004): "*Creativity* masters a process of making or producing, *criticality* a process of assessing or judging" (p. 4). Despite these differences, the relationship between creativity and criticality dispels the notion that creative thinking is vastly different from CT in that the former is "based on irrational or unconscious forces," the latter on "rational and conscious processes" (Paul, 1993, p. 101).

Higher-order Thinking

CT defined by higher-order thinking suggests a hierarchy of cognitive skills in which CT is found at the higher end of the spectrum. This hierarchy is articulated by Bloom who named six levels of cognition from lower- to higher-order thinking, beginning with the more basic and ending with the most complex: Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation (Clark, 2009).

Paul and Scriven's (1996) definition of CT includes terminology similar to that in Bloom's Taxonomy. According to Paul and Scriven, CT is "the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action."

The Scientific Method is an example of a procedure associated with CT because it involves the use of higher-order thinking skills. In this sense, the Scientific Method, a process through which scientists attempt to establish an accurate representation of the world (Wolfs, 1996), cultivates higher-order thinking. The four steps of the Scientific Method: observing and describing a phenomenon; formulating a hypothesis; using the hypotheses to make predictions; and conducting experiments to test the predictions (Wolfs, 1996) suggest the ability to analyze, synthesize, and evaluate information in order to be implemented.

Kurfiss's (1988) definition of CT is in keeping with the Scientific Method in suggesting that it is "an investigation whose purpose it is to explore a situation, phenomenon, question, or problem to arrive at a hypothesis or conclusion about it that integrates all available information and that can therefore be convincingly justified" (p. 2).

CT described in the framework of higher-order thinking in both Bloom's Taxonomy and the Scientific Method is described in a study conducted by the American Psychological Association (APA). Six cognitive skills essential to CT were determined: "Interpretation," "Analysis," "Evaluation," "Inference," "Explanation," and "Self-regulation" (Jackson et al., 2006, p. 35).

Also similar to both Bloom and the Scientific Method is Johnson's (2000) definition that CT "has to do with organizing, analyzing, evaluating, or describing what is already there [and] generally leads the thinker towards a specific conclusion" (p. 46).

As both the APA study and Johnson's findings indicate, definitions of CT that are synonymous with or closely linked to higher-order thinking exhibit a recurring theme. The ability to think critically is expressed primarily as a skill or set of skills with which

one solves problems in a logical, orderly fashion by hypothesizing, experimenting, analyzing, synthesizing, judging, and/or evaluating, in order to arrive at what are considered sound, evidenced-based conclusions.

Critical Thinking Disposition

It is common to find CT defined in terms of skills; however, a more modern approach suggests that skills account for only part CT (Bailin & Siegel, 2003; Brown & Keeley, 1998; Glaser, 1941; Jackson et al., 2006). For example, it has been argued that CT "involves two related, but conceptually distinct, aspects or dimensions: the ability to reason well and the disposition to do so" (Bailin & Siegel, 2003, p. 182).

Glaser (1941) was one of the first theorists to include in his definition of CT the element of disposition, as evidenced in the following quote. According to Glaser, CT is:

> (1) an attitude of being disposed to consider in a thoughtful way the problems and subjects that come within the range of one's experience; (2) knowledge of the methods of logical enquiry and reasoning; and (3) some skill in applying those methods. Critical thinking calls for a persistent effort to examine any belief or supposed form of knowledge in the light of the evidence that supports it and the further conclusions to which it tends. (p. 5)

The phrase "an attitude of being disposed to consider in a thoughtful way" distinguishes Glaser's definition of CT from earlier definitions. The two remaining parts of the definition indicate a link to CT as higher-order thinking in that they include "knowledge," "logical enquiry," "application," "method," "examination," "evidence," and "conclusions."

Like Glaser's definition, Brown and Keeley's (1998) definition of CT includes both skill and disposition. Theirs differs from Glaser's, however, in that it suggests that

CT is a matter of cultivating "skills and attitudes...around a series of critical questions" (p. 2). In this context, CT refers to an "awareness of a set of interrelated critical questions, ability to ask and answer critical questions at appropriate times, and desire to actively use the critical questions" (p. 2). In order to ask, consider, and/or answer critical questions, one would have to be open to the "curiosity, wonder, and intellectual adventure essential to critical thinking" (p. 2).

The ability to formulate critical questions and the disposition necessary to do so are supported by a study conducted by the APA in which seven characteristics fundamental to the CT disposition were determined: "truth seeking," "open-minded," "analytical," "systematic," "self-confident," "inquisitive," and "mature" (Jackson et al., 2006, p. 36).

These disposition qualities characterize the critical thinker as one who purposefully seeks to acquire knowledge; considers arguments from multiple points of view; approaches problem solving in a systematic fashion that encompasses organization, reasoning, and evidence; and is open to reconciling new knowledge with previously held convictions. Jackson et al. (2006) support this claim with their finding of recurring themes in definitions of CT. Their finding combines CT skill and disposition; in this sense, CT requires not only a sturdy foundation of knowledge, but also the willingness to pose questions, and the open-mindedness to consider new possibilities.

Summary of Definitions of Critical Thinking

Although there is not one precise definition of CT, CT is associated with characteristics of reflective, creative, and higher-order thinking. In general, CT involves

reflecting on attitudes and actions (Ennis, 1989; Noddings, 2006; Jackson et al., 2006), creating or considering something in an innovative way (Paul, 1993; Paul & Elder, 2004; Saskatchewan Education, n.d.), and using higher order thinking skills such as Analysis, Synthesis, and Evaluation to draw conclusions with evidence (Kurfiss, 1988; Paul & Scriven, 1996). CT, however, should not be limited to a skill. An important part of the ability to think critically is the disposition to do so (Bailin & Siegal, 2003; Brown & Keeley, 1998; Glaser, 1941; Jackson et al., 2006). The critical disposition includes, for example, being open-minded, analytical, systematic, and inquisitive (Jackson et al., 2006).

Critical Thinking Instruction

Studies indicate that there is a greater chance of developing students' CT skills when instruction includes a deliberate effort to teach CT (Ennis, 1989). Ennis's (1989) "Mixed Approach" has been found effective (Barak, Ben-Chaim & Zoller, 2007; Finlay & Faulkner, 2005; Lee, 2006; Plath, English, Connors & Beveridge, 1999; Sokol, Oget, Sonntag & Khomenko, 2008). In particular, instruction using the Mixed Approach combined with Socratic Questioning, connections to the real world, and thinking in terms of multiple perspectives and possibilities, has been found to promote CT abilities (Brady, 2008; Dong, 2006; Hannel & Hannel, 1998; Lampert, 2006; ten Dam & Volman, 2004).

Several studies indicate that when the Mixed Approach (Ennis, 1989) was implemented, CT skills were developed more effectively than in traditional teaching situations (Barak, Ben-Chaim & Zoller, 2007; Finlay & Faulkner, 2005; Lee, 2006; Plath et al., 1999; Sokol, Oget, Sonntag & Khomenko, 2008). This approach is considered effective because it includes the deliberate teaching of CT (Ennis, 1989). The Mixed

Approach involves either "infusion," wherein the principles of CT skill and dispositions are explicit, or "immersion," wherein they are implicit (Ennis, 1989).

Mixed Approach to Teaching Critical Thinking

The Mixed Approach to teaching CT is a combination of the General Approach with either Immersion or Infusion (Ennis, 1989).

The General Approach is based on investigating relationships between variables, such as (A x B) = (B x A) in mathematics; it suggests a universal set of CT skills applicable across curriculum and to any given situation where CT is required (Ennis, 1989). The General Approach has been criticized on several grounds: that it focuses on abstract variables rather than on true CT and relevant content (Willingham, 2007); suggests CT skills as technical skills, rather than skills within content area domain (ten Dam and Volman, 2004); is "formulaic" and may limit students to criticizing the arguments of others, without teaching them to develop and support their own positions (Noddings, 2006, p.33); and it focuses more on learning about CT, than engaging in CT itself (van Gelder, 2005). The General Approach is of limited use on its own (Plath et al., 1999).

The Mixed Approach to teaching CT combines the General Approach with either Infusion or Immersion. Like the General Approach, the Mixed Approach recognizes that CT skills may be applied to different academic domains and situations. It differs in that it stresses that CT thinking must be taught either explicitly (Infusion) or implicitly (Immersion) in the context of relevant subject matter (Ennis, 1989). While the General

Approach by itself has limited use, the Mixed Approach provides students opportunities to identify, apply, and transfer skills in different contexts (Plath et al., 1999).

Teaching CT in the context of personal beliefs and decision-making supports the idea that "students learn best to think critically by hearing, thinking, and talking about issues critical to their present or future lives" (Noddings, 2006, p. 32). A similar argument has been made by Kuhn (1999), who exerts that "we cannot expect children to progress in the development of thinking unless we give them something to think about, in other words, unless we engage them in serious learning about meaningful, rich, domain-specific subject matter" (p. 17).

Studies indicate that the Mixed Approach, using either Infusion or Immersion, increases CT skills and disposition (Barak, Ben-Chaim & Zoller, 2007; Finlay & Faulkner, 2005; Lee, 2006; Plath et al., 1999; Sokol, Oget, Sonntag & Khomenko, 2008). Unlike more traditional instruction, or instruction in which the teaching of CT is not deliberate, the Mixed Approach scaffolds CT strategies for students (Willingham, 2007) and develops habits of mind that foster long-term CT skills (Barak et al., 2007).

Infusion

Infusion is an approach to teaching CT that is, according to Ennis (1989), "deep, thoughtful, well understood subject-matter instruction in which students are encouraged to think critically in the subject, and in which general principles of critical thinking dispositions and abilities are *made explicit*" (p. 5).

Results of numerous studies attest to the effectiveness of this approach (Barak, et al., 2007; Plath et al., 1999; Sokol, et al., 2008). One study determined that science

students who were taught by teachers using strategies to intentionally enhance CT skills and disposition showed significantly greater improvement on post-tests of CT than students in control groups (Barak, et al., 2007). Researchers concluded that "a persistent effort to promote higher-order thinking not only fosters students' critical thinking capabilities during the learning period, but also contributes to the sustainability of these capabilities in the long run, as an integral part of students' thinking habit" (Barak et al., 2007, p. 360).

A study conducted in a social work program found that the group of participants who were instructed using the Infusion Approach demonstrated significantly higher achievement between pre- and post-test scores on tests of CT skill and disposition than those in control groups (Plath et al. 1999). The improvement of their CT skills was attributed to the idea that the explicit instruction of CT in the Infusion Approach develops more complex CT skills and a deeper awareness of their application (Plath et al., 1999).

In another study, foreign language students in the experimental group showed significant gains in inventive thinking skills in the context of the target language compared to those in the control group (Sokol, et al., 2008). The experimental group was better equipped with the CT skills necessary to learn target language grammar; to apply the target language as a means for solving problems; to understand how various problem-solving models work; and to transfer knowledge and thinking skills to new contexts. Researchers conclude that this was due to their exposure to explicit CT instruction geared toward solving complex problems (Sokol et al., 2008).

Immersion

Immersion, according to Ennis (1989), is a similar to Infusion in that it is "thought-provoking...subject-matter instruction in which students do get deeply immersed in the subject, but in which general critical thinking principles are *not made explicit*" (p. 5). Through Immersion, the acquisition of CT skills is implicit; it is similar to the acquisition of a second language as learned in an immersion program wherein students are "immersed" in learning the language by listening and speaking uniquely in the target language, without receiving explicit instruction of the language's mechanics.

Lee's (2006) "moral engagement approach" is an example of Immersion used in a course on ethics. Rather than receive explicit instruction of CT, students were taught to think critically by participating in class conversations about euthanasia, capital punishment, or other issues of debate. Lee suggests that through class conversations, rather than direct instruction, students improved their ability to "listen to others in an open-minded manner and com[e] to carefully considered conclusions only after thoughtful reflection about differing views concerning matters of controversy" (p. 199). Likewise, some of the criteria for assessing quantitatively students' written work, were "conceptual clarity," "supporting reasons for position taken," "identification of opposing arguments," and "responses to opposing arguments" (p. 204). Through the Immersion Approach, the ability to think critically, as evidenced by successfully meeting these criteria, is acquired by participating in class activities in which "students develop their listening skills [and] delineate and respond to opposing arguments" (p. 205).

Finlay and Faulkner's (2005) study is another example of the Immersion Approach. In this study, they sought to improve students' CT abilities and level of

engagement in reading through peer learning. Peer learning refers to the implementation of instructional and learning strategies in which students learn from one another with limited participation from the teacher (Boud, Cohen & Sampson, 1999). It encourages CT through greater content understanding, refined judging and interpretation skills, and increased awareness of one's own thinking (Finlay & Faulkner, 2005). In addition, it entails changing the focus from what is taught, to what is learned, and requires that the student rather than the teacher be responsible for the acquisition, organization, and application of new knowledge (Marie & Cooper, 2002).

In this particular study, traditional lecture-based instruction was replaced with peer learning groups of 3 to 5 students in each. Students summarized readings, shared synopses, engaged in discussions led by guiding questions, and made informal presentations of the results of their discussions, in which they compared and contrasted common themes. Results of the study showed greater comprehension, higher levels of participation, and more active learning, all of which suggests that the peer reading groups encouraged a process of CT (Finlay & Faulkner, 2005).

Characteristics of Critical Thinking Instruction

Deliberate instruction most likely to foster CT, regardless of subject area, is characterized primarily by three components: Socratic questioning; making connections to "real life"; and thinking in terms of multiple perspectives and possibilities (Brady, 2008; Dong, 2006; Hannel & Hannel, 1998; Lampert, 2006; ten Dam & Volman, 2004). These characteristics refer respectively to problem solving through open-ended questioning; applying what is learned in the context of academics to situations outside the

classroom; and considering different points of view and accepting that a solution need not come from any single "right" answer. These characteristics of instruction develop CT skills and lend themselves to open-mindedness, creativity, inquisition, curiosity, and reflection (Brady, 2008; Dong, 2006; Hannel & Hannel, 1998; Lampert, 2006; ten Dam & Volman, 2004).

Socratic Questioning

To teach CT, instruction should incorporate systematic Socratic questioning over teacher-led lectures or direct instruction (Hannel & Hannel, 1998; Noddings, 2008). Unlike factual memorization, open-ended questions require thought and analysis (Noddings, 2008). Socratic questioning requires students to sift through information, consider it from different perspectives, and form opinions with sound evidence (Hannel & Hannel, 1998). It has been found that teaching students to formulate questions through explicit questioning techniques provides opportunities to hypothesize, reflect, and examine information, which promotes CT and high-level responses (Dong, 2006).

Open-ended, question-based activities develop CT skills by encouraging students to consider others' responses and perspectives and reconcile old knowledge with new ideas (Lambert, 2006). It also encourages students to connect with the world around them (Dong, 2006).

Connections to Real Life

Encouraging students to make connections between academic material and the world outside the classroom is another strategy for teaching CT (Hannel & Hannel, 1998;

Dong 2006; Brady, 2008; Noddings, 2008; ten Dam & Volman, 2004). By replacing instruction that focuses on rote learning and recall with instruction that focuses on innovative thinking about relevant, real world challenges, requires students to draw from every known thinking skill to make sense of their daily experiences (Brady, 2008).

According to ten Dam and Volman (2004) "critical competence" is the skill born of the connection between academic content and the real world, and it is this skill that stimulates students to actively participate in the betterment of society. This idea is supported by Brady (2008) who contends that the most effective way to teach CT is to "focus attention directly on some part of the real world" (p. 67).

Multiple Perspectives and Possibilities

Instruction that incorporates open-ended questioning and connections to the real world encourages CT by presenting students with multiple perspectives and possibilities, as opposed to one "right" answer (Lampert, 2006). Group conversations wherein possibilities are discussed encourage students to consider different answers and reconcile various and opposing viewpoints (Lampert, 2006).

To engage in conversations in which possibilities are explored and differing perspectives are valued, an open-minded critical disposition is essential. ten Dam and Volman (2004) explain:

> [Critical thinking] is not only about issues such as 'being able to look at a problem from different perspectives (the ability to change perspective)' or 'being able to pinpoint the main issues and assumptions in a discussion or argument'. It also concerns skills and dispositions like 'being able to relate a question to one's own norms and values', 'being able to relate a question to general principles such as social justice, equality, respect and consideration', 'be open to and consider other people', 'daring to express a different opinion', etc. (p. 373)

Instruction of this sort challenges students intellectually and "enables students and teachers to experience the 'meatiness' of the direct study of reality.... It shifts the emphasis from cover-the-material memory work to a full range of thinking skills" (Brady, 2008, p. 67).

Summary of Critical Thinking Instruction

Void of context, the General Approach to CT instruction is hardly effective (Noddings, 2006; ten Dam & Volman, 2004; van Gelder, 2005; Willingham, 2007). As studies indicate, CT is most likely to occur when taught from an approach that blends the teaching of general CT skills with contextualized subject matter (Barak et al., 2007; Plath et al., 1999; Sokol et al., 2008). Instruction that includes the deliberate, contextualized teaching of CT, be it explicit or implicit, is most likely to develop students' CT skills (Barak et al., 2007; Finlay & Faulkner, 2005; Kuhn, 1999; Lee, 2006; Plath et al., 1999; Sokol et al., 2008).

Several studies indicate the effectiveness of the Mixed Approach because it teaches CT within academic context and builds skills that transfer across disciplines and apply to different circumstances (Barak et al., 2007; Dong, 2006; Finlay & Faulkner, 2005; Lampert, 2006; Plath et al., 1999; Sokol et al., 2008; ten Dam & Volman, 2004; van Gelder, 2005; Willingham, 2007). This approach, combined with Socratic questioning, the making connections between academic content and the real world, and consideration of information from multiple perspectives, develops students' CT skills (Brady, 2008; Dong, 2006; Hannel & Hannel, 1998; Lampert, 2006; ten Dam & Volman, 2004). It has been found that "...if teachers purposely and persistently practice higher

order thinking strategies, for example, dealing in class with real-world problems, encouraging open-ended class discussions, and fostering inquiry-oriented experiments, there is a good chance for a consequent development of critical thinking capabilities" (Barak et al., 2007, p. 353).

Critical Thinking Assessment

"Multi-aspect tests," also referred to as "general tests" of CT, assess a variety of CT skills; "aspect-specific tests" are designed to assess one area of CT (Ennis, 1999). Multi-aspect assessment is explored here because it is the type of test that was used in this study.

Multiple-choice Tests

The Watson-Glaser Critical Thinking Appraisal (WGCTA) is one of the more popular tests used to assess the CT abilities of high school and college aged students (Ennis, 1999). It is reputed for its validity and reliability (El Hassan & Madhum, 2007).

That said, multiple-choice test of CT such as the WGCTA have received criticism that they are too limiting to make an accurate assessment of CT skills. For example, one section of the WGCTA provides only two multiple-choice options; Wagner and Harvey (2006) speculate that success on this part of the test could be attributed to successful guessing rather than the true application of CT skills. In addition, multiple-choice tests do not allow for individual responses; they may ask students to select a hypothesis, generalization, value judgment, or another higher-order thought process from a list of options, but "...guessing what an anonymous writer of a test item had in mind isn't

higher-order thinking" (Brady, 2008, p. 67). Brady (2008) questions the validity of multiple-choice exams because they offer only one correct response, as designated by the makers of the test. For example, the WGCTA has been questioned because it "...measures the ability of persons to follow the 'rules' involved in various forms of reasoning. To the extent that one can accept the underlying rules, the test is valid" (p. 20).

Because of the limitations of multiple-choice tests, some studies that employed multiple-choice tests as their primary means of assessment supplemented them with subjective testing methods such as essays or informal interviews to yield more thorough results (ten Dam & Volman, 2004).

Essay Tests

The advantages of assessment in essay-format are several. Unlike multiple-choice exams where answers are provided, essay tests require one to formulate his own response, which suggests that essay tests evaluate higher-level thinking more thoroughly than multiple-choice tests (Oermann, 1999; Tomey, 1999; Werner, 1991) and yield individual responses as opposed to uniform answers (Brady, 2008). Responses to essay tests require analysis, evaluation, argument, and evidence, and therefore assess a more complicated thought process than those of multiple-choice tests (Brady, 2008; Brunt, 2005).

The Ennis-Weir Critical Thinking Essay Test (EW)

The EW is a popular instrument in studies of CT (Werner, 1991). It includes the components that make up a sound CT test design in that it provides multiple possibilities for answers, room for varied ways of looking at problems, leeway for changing opinions

based on changing circumstances, and opportunity to apply knowledge in a relevant context (Ennis & Weir, 1985; Werner, 1991; Yeh, 2001). The EW tests general CT skills in the context of argumentation (Ennis & Weir, 1985). It is intended to evaluate a person's ability to assess an argument and to formulate an argument in response, thus suggesting a creative dimension in CT (Ennis & Weir, 1985).

The EW has been criticized on the grounds that its content (i.e., overnight street parking) may present a dilemma for students who are not familiar with the issues surrounding parking, driving, and rush hours (Ennis & Weir, 1985). The fact that the EW does not include two different versions for pre- and post-tests could be considered a drawback. Studies that incorporate the EW in a pre-/post-test design administer the same test twice; therefore, improvement in scores could be attributed to testing effect, rather than acquired skill.

Summary of Critical Thinking Assessment

Although some multiple-choice tests of CT are highly regarded instruments, they cannot include the design, skill requirement, subjectivity, individuality, and contextualization provided by essay tests such as the EW (Werner, 1991).

Mirroring the principles of CT instruction (i.e., Socratic Questioning, connections to real world experiences, and possibilities), the EW requires that one question and argue in an effort to problem-solve within a relevant, real life context. The subjectivity of the test's design encourages students to think in terms of possibilities and consider information from more than one stance. Ennis and Weir explain:

> This is not a test of formal or deductive argument, nor does it require technical knowledge of such. In comparison with arguments considered in

> many deductive logic tests, arguments in the real world require considerable interpretation (in context), require evaluation of content as well as form, often have value dimensions, and do not have mechanical decision procedures. This is a real-world test. (p. 1)

Conclusions

Although a collective definition of CT continues to elude theorists because of the complex nature of thinking (Paul, 1993), there is evidence of overlap among definitions described in terms of "reflective thinking," "creative thinking," and "higher-order thinking" (Noddings, 2006; Paul & Elder, 2004; Kurfiss, 1988).

Ennis's (1989) Mixed Approach is considered an effective way to teach CT, as evidenced by studies in which students taught using the Mixed Approach showed significantly higher levels of CT ability than students in control groups (Barak et al., 2007; Finlay & Faulkner, 2005; Lee, 2006; Plath et al., 1999; Sokol et al, 2008). To teach CT effectively, the Mixed Approach should include Socratic Questioning, connections to the real world, and multiple perspectives and possibilities (Brady, 2008; Dong, 2006; Hannel & Hannel, 1998; Lampert, 2006; ten Dam & Volman, 2004).

CT assessment should align with CT definitions and instruction in order to evaluate CT skills most effectively. *The Ennis-Weir Critical Thinking Essay Test* (EW) is an example of sound alignment because it encourages questioning, is contextualized in relevant content matter, provides opportunity for multiple interpretations and different answers, and requires that students apply CT skills to formulate responses unique to their analysis and interpretation.

CHAPTER THREE: METHOD

Overview

This chapter presents the study in five parts. The first part describes the schools involved in the study. The second part describes the procedures used to obtain participants, and the demographics of each group. The third part examines the assessment instrument. The fourth outlines the experimental design and data collection procedures. The fifth identifies procedures for analysis.

Methodology

Setting

This study took place at two schools in the same mid-sized city in the northwestern United States. One of the schools was selected because it offered the International Baccalaureate (IB) program. The other school was chosen because it offered the option to take one or more Advanced Placement (AP) classes, or none at all. The students in this study were in Grade 12. Both of the schools in this study were private. The test was administered and data collected toward the end of the third quarter of the school year. Site A's school year ran from late August to early June. Site B's school year ran from mid-August to late May.

Site A

Site A was a coeducational, independent, K-12 school with a vision of academic excellence, innovative thinking, and creative self-expression. All Grade 11 and 12 students at this school were enrolled in the IB program. They were required to take a philosophy course and six IB classes: native language, second language, individuals and societies, experimental sciences, math/computer science, and the arts. In addition, students were required to complete a senior research project and fulfill a 150-hour requirement of creativity, action, and service. Three hundred thirty students were enrolled in Grades K-12, with 111 in Grades 9-12. The student to teacher ratio in Grades 9-12 was 15:1.

The tuition to attend this school is $11,000 per year in Grades 9-12. Thirty-nine percent of the school population receives some form of financial aid. The school offers financial assistance by way of merit-based and need-based scholarships. Awards are determined by the admissions committee.

Site B

Site B was a coeducational, private, Catholic, 9-12 school with a vision of academic excellence, moral leadership, and social responsibility in allegiance to the Catholic church. This school offers AP classes in art, Calculus, chemistry, physics, English, Spanish, statistics, and United States history. Approximately 680 students were enrolled. The student to teacher ratio was 18:1.

The tuition to attend this school is $8,000 per year for international/exchange students; $6,500 for students not registered with a local parish; and $5,700 for students

who are registered with a local parish. School B offers three types of financial assistance: "Tiered Tuition," reduced tuition rates for families with more than one child attending the school at the same time (10% reduction for second child; 20% reduction for third child; 30% reduction for fourth child); "Financial Grant" for families who qualify (< 100% tuition paid); and "Work Study" for students who contract to work for the school during the summer months in exchange for tuition.

Participants

Three groups of 20 students from two different schools participated in the study. All were in Grade 12 and enrolled in the IB program at Site A, or the AP or non-AP (NON) program at Site B. The number of participants per group was set at 20 because that was the maximum number of IB participants available for the study. AP and NON groups were matched with the IB group according to number of participants, gender, native language, parental education level, family income, SAT scores, years of attendance at current school (Grades 9-12); and grade point average (GPA). This measure was taken to equate groups, thus eliminating the alternative explanation that differences in test scores across groups could be attributed to demographic differences (Gay, Mills & Airasian, 2008).

International Baccalaureate (IB) Students

The researcher contacted the principal at Site A (see Appendix A for letter). Two IB English teachers agreed to have their students recruited for the study. The researcher recruited participants in three sections of IB English. In each class, the study was

described, assent and consent forms were distributed, and questions from students were addressed (see Appendix B for script; Appendix C for assent forms; and Appendix D for consent forms). As an incentive, students who submitted signed forms to the school receptionist by a certain date were given a gift card, regardless of whether or not they consented to participate in the study.

Advanced Placement (AP) Students

The researcher contacted the vice principal at Site B regarding her study (see Appendix A for letter). A teacher with whom the researcher, had worked with previously in a capacity not related to school, enquired at Site B if there were teachers who would be interested in participating in the study. A government teacher of both AP and non-AP Grade 12 students at Site B volunteered his classes to be recruited for the study. His students were recruited using the same script and assent and consent forms used at Site A (Appendices B, C, and D). At the teacher's discretion, students received 10 class points for returning their permission slips, even if they declined to participate in the study. As an incentive, students who returned submitted signed forms to their government teacher by a certain date were given a gift card, regardless of whether or not they consented to participate in the study.

From a pool of 118 students who took the test at Site B, 20 AP students were selected who shared the same demographics as the IB students. The results of tests of students who did not fit the criteria were not included in the study.

Students Not in a Specific Program (NON)

NON students were students of the same government teacher. They were recruited in their government class with the same script (Appendix B) and in the same manner as the AP students at Site B.

Of the 118 students who took the test at Site B, 20 NON students were selected who had demographic profiles similar to the IB group. Tests of unmatched students were not analyzed.

Selection Procedures

The total number of IB participants was 20 (8 males; 12 females). There were no non-IB students to recruit because all Grade 11 and 12 students at Site A are automatically enrolled in the IB program.

The total number of AP and NON participants was 118. To equate groups by number of participants and demographics, 20 AP and 20 NON students who matched the IB students' characteristics of gender, native language, parental education level, family income, SAT scores, years at current school, and GPA were selected. Tests of students who did not match the IB students demographically were not analyzed for the study.

Demographic Characteristics

Table 3.1 indicates that gender, native language, parental education, and family income was close to equivalent across groups. Table 3.2 indicates that SAT scores and years at current school were close to equivalent; however, there was a significant difference in group GPA.

IB, AP, and NON were close to equivalent in terms of categorical demographics. The breakdown of groups by gender, native language, parental education, and family income is indicated in Table 3.1.

To verify the comparability of groups, a series of chi-square analyses was conducted. Gender was not related to program, $X^2(2) = .000$, $p = 1.0$, which indicates that the gender breakdown did not differ across groups. Native language was not related to program, $X^2(4) = 2.111$, $p = .715$, which indicates that native language did not differ across groups. Parental education was not related to program, $X^2(3) = 4.185$, $p = .381$, which indicates that the parental education levels did not differ across groups. Family income was not related to program, $X^2(6) = .645$, $p = .996$, which indicates that family income did not differ across groups.

Table 3.1: Breakdown of Categorical Variables across Groups.

	IB	**AP**	**NON**	**p**
Gender				1.0
Males	8	8	8	
Females	12	12	12	
Native Language				.715
English	18	17	19	
Bilingual	1	2	0	
Non-English	1	1	1	
Parental Education				.381
High School	3	3	1	
Bachelor's	1	1	4	
Graduate	16	16	15	
Family Income				.996
Less than 50K	2	1	1	
50K – 99K	1	1	1	
100K – 200K	2	3	2	
Greater than 200K	9	8	6	

IB, AP, and NON were close to equivalent in terms of continuous variables, except for GPA. The means (with standard deviations in parentheses) of average SAT scores, number of years at current school (Grades 9-12), and GPA are indicated in Table 3.2. The asterisk indicates a significant difference across groups.

To verify the comparability of groups, a series of one-way analysis of variance (ANOVA) with an alpha level of .05 was conducted based on continuous variables. SAT did not differ across the three groups, $F(2, 27) = .588$, $MSE = 59688.37$, $p = .563$, which indicates that mean SAT did not differ significantly across the three groups. Years of enrollment (in Grades 9-12) did not differ across the three groups, $F(2, 58) = .677$, $MSE = .628$, $p = .512$, which means that the number of years enrolled in Grades 9-12 at one's current school did not differ significantly across groups.

Table 3.2: Breakdown of Continuous Variables across Groups.

	IB	**AP**	**NON**
Mean SAT	1734 (268.40)	1761 (250.20)	1628 (149.41)
Years at current School (9-12)	3.6 (.840)	3.6 (.994)	3.6 (.402)
Mean GPA*	3.52 (.363)	3.78 (.190)	3.16 (.375)

GPA was differed across groups, $F(2, 53) = 16.698$, $MSE = .105$, $p < .001$. A follow-up Tukey HSD indicates that IB and AP had significantly higher cumulative GPA than NON. This was the only significant difference found between groups; however, it was not considered a serious confound since several studies show that cumulative GPA and scores on tests of CT are unrelated (Conger & Mezza, 1996; Cox, 2008; Leaver-

Dunn, Harrelson, Martin & Wyatt, 2002; Scott, Markert & Dunn, 1998; Williams, Glasnapp, Tilliss, Osborn, Wilkens, Mitchell, Kershbaum & Schmidt, 2003).

Instrument

The Ennis-Weir Critical Thinking Essay Test (EW) was the instrument used to measure participants' CT skills. It was selected from a number of commercial tests of CT because it is an essay test and is appropriate for use with students in Grade 7 through college (Ennis & Weir, 1985). In addition, open-ended tests of CT such as the EW envelop the principles of CT in that they have been found to encourage application of CT skills, elicit multiple interpretations, and engender possibilities in the context of solving real life problems (Brady, 2008; Oermann, 1999; ten Dam & Volman, 2004; Tomey, 1999; Yeh, 2001).

Participants read a fictitious letter to the editor consisting of eight paragraphs (see Appendix E for test). In writing a letter in response, participants addressed the argument(s) presented in each paragraph of the original letter, followed by a concluding paragraph. In keeping with the EW scoring rubric, students received -1 to 3 points for their first eight paragraphs, and -1 to 5 points for their conclusion (see Appendix F for rubric).

Design and Procedures

This quantitative study included three groups of students in different academic programs. It included one test of CT, the EW, with outcomes used as variables in statistical analyses. A series of one-way ANOVA was run to compare group averages.

The alpha level was set at .05. The independent variable was the academic program of enrollment: IB, AP, or NON. The dependent variables were the eight individual paragraphs of the EW as they correspond to five CT subscales: Evaluation of Argument (Paragraphs 1, 2, and 5); Deduction (Paragraphs 3 and 8); Inference (Paragraph 4); Recognition of Assumption (Paragraph 6); and Interpretation (Paragraph 7).

The subscales were determined by a review of CT assessments: The Cornell Critical Thinking Test; the California Critical Thinking Skills Test; the EW; and the Watson-Glaser Critical Thinking Appraisal. Each of these tests assesses the ability to evaluate arguments, deduce, infer, recognize when assumptions have been made, and interpret information (Ennis & Weir, 1985; Ennis, Millman & Tomko, 2005; Pike, 1997; Watson & Glaser, 2006).

Evaluation of Argument

Evaluation of Argument is defined as the ability to distinguish between strong and weak arguments (Watson & Glaser, 2006). The researcher assigned Paragraphs 1, 2, and 5 to the Evaluation of Argument subscale because they test the ability to pinpoint the weaknesses of each argument (see Appendices E and F for EW test and rubric).

Deduction

Deduction is the ability to extract information in order to form a sound conclusion (Ennis, et al., 2005). The researcher assigned Paragraphs 3 and 8 to the Deduction subscale because they are the only two paragraphs in the letter that are considered sound (see Appendices E and F).

Inference

Inference pertains to a conclusion that is made based on observation or supposition of fact (Watson & Glaser, 2006). The researcher assigned Paragraph 4 to the Inference subscale because the paragraph questions whether the conclusion to ban overnight parking can be made based on the author's conviction that it is "bad" and "undesirable" (see Appendices E and F).

Recognition of Assumption

Recognition of Assumption pertains to the ability to recognize when a conclusion has been made without adequate information (Ennis et al., 2005). The researcher assigned Paragraph 6 to the Recognition of Assumption subscale because the author's conclusion is drawn based on the results of a poorly constructed experiment that does not provide sufficient information or evidence to fully support the author's claim (see Appendices E and F).

Interpretation

Interpretation involves judging whether logical conclusions can be drawn with certainty based on given information (Watson & Glaser, 2006). The researcher assigned Paragraph 7 to the Interpretation subscale because the author's language must be interpreted in order to uncover the limitation of his subjective, incorrect definition a term (see Appendices E and F for EW).

Demographic Data Collection

Each participant completed a voluntary, paper-and-pencil demographic questionnaire, created by the researcher, on gender, grade level, years of attendance at current school, academic program of enrollment, native language, parental education level, zip code, self-reported GPA, self-reported standardized test scores (PSAT, SAT, and ACT), and estimated family income (see Appendix G for questionnaire).

Site A students completed the questionnaire at the same time as the test. Site B students completed the questionnaire the week before the test; their teacher collected the questionnaires and redistributed them to students on the day of the test.

Test Administration

Participants took one test, administered by the researcher, in an approximately 50-minute class period during a school day in the third quarter of their Grade 12 year.

IB students took the test during their English classes. AP and NON students took the test during their government class. In both cases, the researcher and the classroom teacher were present for the test; students were seated at desks and the room was quiet. Each participant was provided a paper-and-pencil testing prompt, including directions, and two pieces of blank, lined paper.

The researcher instructed students to read the directions of the EW to themselves as she read it aloud to the class (Appendix E). The only change that was made to these instructions was in the second to last paragraph: "Sign your name to your letter. You are a local citizen, and this topic concerns you." Students were instructed to sign a fake name. This was done to ensure anonymity of participants.

Students had 40 minutes to complete the test. Those who finished before the allotted time read or did schoolwork and remained quiet for the rest of the period, as requested by their classroom teacher.

Upon completion of the test, students submitted their essays, along with the test directions, test prompt, and demographic questionnaire. Papers were stapled together in that order by the researcher. This was done so that demographic questionnaires were not visible to assessors. Tests were scored blindly, thus minimizing any potential influence that knowledge of demographics might have on scoring (Campbell & Stanley, 1963) and any bias the researcher may have otherwise had due to her former position as a teacher at Site A.

Assessment Procedures

Sixty essay tests of CT were scored. In order to do so in an accurate and consistent manner, the researcher established inter-rater reliability through a three-step process: Setting Criteria; Scoring Tests; and Reconciling Results.

Morris, Fitz-Gibbon & Lindheim (1987) suggest that the most effective way to reduce inconsistency among human instruments is to have more than one person score the tests. To establish reliability of essay tests in particular, it is best to train another person to score in the same manner as the principal reader and have him/her score a sample of the tests (Morris et al., 1987). Calculating the scores awarded by both assessors provides an estimate of the extent to which the principal reader can be relied upon to consistently score the tests with accuracy (Morris et al., 1987).

The EW has an inter-rater reliability measure of 82 to 86 percent, which is high for an essay test (Ennis & Weir, 1985). One hundred percent of tests included in the study were assessed by the researcher (N = 60). To establish inter-rater reliability, she trained a colleague from the same doctoral program to score the EW. This colleague also assessed 100 percent of the tests.

The rate of inter-rater reliability of scoring for this study was high. With nine paragraphs in each of 60 tests, there were 540 individual paragraphs to score. In the end, 532 paragraphs (98.5 percent) had the same score; eight paragraphs (1.5 percent) differed by one point. These differences were found in Paragraphs 1, 5, and 9. In terms of total test scores, 86 percent of exams (N = 52) had the same score for each paragraph; 14 percent of exams (N = 8) differed by one point.

Prior to scoring the tests, examiners met to discuss the specifics of scoring, including the elements necessary to obtain full credit, partial credit, no credit, or a score of -1. The first eight paragraphs were scored on a scale of -1 to 3. In keeping with the EW scoring rubric, +3 was awarded for adequate justification; +2 for semi-adequate justification; +1 for correct judgment without justification; 0 for lack of judgment and no reason(s) given; and -1 for poor judgment in justification or incorrect judgment (Ennis & Weir, 1985) (Appendix F).

The ninth paragraph was scored on a scale of -1 to 5.[1] Students were awarded one point for denouncing the argument as a whole; one point for summarizing points made in the preceding eight paragraphs; two points for addressing the error of drawing

---

[1] Because Paragraph 9 was a summary, it was not assigned to a particular subscale. Scores on Paragraph 9 were included only in students' final scores on the EW.

conclusions about all streets based on only busy streets, and one point for noting the author's use of emotive language to persuade his audience (Ennis & Weir, 1985).

After scoring two tests together, the examiners felt they understood the scoring criteria well enough to be consistent scoring the remaining exams separately.

Separately, examiners both scored the remaining 58 exams. All tests were scored blindly and in keeping with the EW scoring rubric. Each test required approximately 15-20 minutes to score. The first eight of the nine paragraphs in students' essays was scored on a scale of -1 to 3 points; the concluding paragraph which was scored on a scale of -1 to 5 points because it was a summary and required more information (Ennis & Weir, 1985).

After scoring the remaining exams, examiners met for four hours on three separate days to discuss scores. They found that because of the scoring possibilities in each of the nine paragraphs, total scores sometimes differed by as much as 14 points. Some tests with the same final score showed variability in scores per paragraph.

The examiners reevaluated the exams, discussed the reasoning behind their scores, carefully reviewed the assessment criteria, and reconciled scores based on systematic, agreed upon negotiation. Upon secondary review, some paragraphs received higher scores as one examiner was able to point out ideas in the writing that may not have been apparent in the initial assessment; in some cases, an examiner realized she had been too generous and therefore reissued a lower score to make a more accurate assessment.

Data Analysis

Two essential questions guided the study:

1. What differences in critical thinking skills, if any, exist between students enrolled in academic programs that purport to foster critical thinking, namely the IB and AP programs, and students in a control group?
2. What differences in critical thinking skills, if any, exist between students enrolled in the IB or AP program, both of which purport to foster critical thinking?

To seek answers to these questions, an empirical study was designed to measure quantitatively the CT skills of students enrolled in the IB program, AP classes, or neither program (NON). Group averages of five CT subscales were compared: Evaluation of Argument; Deduction; Inference; Recognition of Assumption; and Interpretation.

Examples of student work are provided to demonstrate how scores were assigned for each subscale. The first examples are of work that received a high score of 2 or 3, the second of work that received a low score of -1, 0, or 1.

Evaluation of Argument

Evaluation of Argument was based on the sum of scores on Paragraphs 1, 2, and 5, all of which present flawed arguments (Ennis & Weir, 1985). Work that highlighted these flaws received high scores. For example, in Paragraph 1, one student identified the author's misuse of analogy by pointing out the weakness of equating overnight parking to having a garage in the streets. He/She wrote:

> Raywift's first argument was not a strong one. Parking overnight in the streets does not imply having a garage in the streets. It would be ridiculous

> to see a garage in the middle of the road (and this would be illegal) but parking in the street, where it is clear that the car owner does not own the street, is not ridiculous and does not imply having a garage in the street.

With regard to Paragraph 2, the same student noted the irrelevance of the argument because afternoon rush hour traffic does not relate to the problems of overnight parking. He/She wrote, "Raywift's argument is not strong either because he does not address the problem at hand. The issue is overnight parking, not afternoon parking, so when Raywift refers to 'afternoon rush hour' this makes the argument completely irrelavent (sp?)."

The same student found in Paragraph 5 the flaw in the claim that eliminating accidents between parked and moving vehicles will affect (i.e., nearly eliminate) other accidents. He/She wrote:

> There are a few problems with Raywift's fifth argument. Although a ban on overnight parking would nearly eliminate accidents between parked and moving vehicles, R. then says that there would be a "near elimination of accidents," and he assumes that if this is "highly desirable" then "intelligent citizens" will give up their overnight parking rights.

Another student had a different stance on these arguments. In response to Paragraph 1, he/she wrote, "I would agree that if it is against the law to have a garage in the city streets then it clearly shouldn't be done. There are no exceptions to laws." In response to Paragraph 2, he/she wrote, "I would agree with reason two. Instead of removing cars from the streets I propose expanding the width of the street. This too would give people more room and still maintain a parking system." In response to Paragraph 5, he/she wrote, "I would have to agree with this statement. Eliminating accidents on main streets would make traffic flow easier. There is also less danger in driving down these roads if accidents can be eliminated." Points were deducted from

Paragraph 1 because the misuse of analogy and/or of an incorrect definition was not noted, from Paragraph 2 because the irrelevance of the argument was not recognized, and from Paragraph 5 because other possibilities were not considered.

Deduction

Deduction was based on the sum of scores on Paragraphs 3 and 8. Participants who identified these paragraphs as sound (Ennis & Weir, 1985) received high scores. With regard to Paragraph 3, one student wrote:

> I believe this argument is coherent and well supported. It appears as though the parking in the streets is affecting the time schedule of a workforce, in which case it should be dealt with. Based on this argument I would agree that parking in the streets should be prohibited.

For Paragraph 8, the same student wrote, "I consider this argument sound because it is supported by expert opinion. Based on this argument I would agree with Mr. Raywift's opinion."

If participants disagreed with the logic in Paragraphs 3 and/or 8, they had to counter with a well-constructed argument in order to receive high scores. Examples of work in which students disagreed and/or contradicted the logic in the paragraph, but could not propose a counter argument strong enough to earn high marks, include, in response to Paragraph 3, "Since the factory workers proceed to work at an early time, they shouldn't have a problem in that there is no traffic there in the morning. Parked cars shouldn't be a problem." and in response to Paragraph 8:

> This is a cyclical argument with no coherent conclusion. By stating that "prohibiting parking from 2 am to 6 am is the best way to prevent overnight parking" the reader my [sic] be confused as this argument is compleate [sic] common sense. If you prohibit overnight parking than

> [sic] overnight parking will be prevented. This argument has no point and is unnecessary in this letter.

In Paragraph 3, points were deducted for not addressing the point that was made. In Paragraph 8, points were deducted for suggesting that it is obvious that prohibiting overnight parking will prevent people from parking on the street. Prohibiting something is meant to discourage it from happening (Ennis & Weir, 1985). Contrary to what the student suggested, the two terms are not analogous; it does not follow that because something is prohibited, it will necessarily be prevented (Ennis & Weir, 1985).

Inference

Inference was based on scores on Paragraph 4. Participants who recognized circularity of language and/or a lack of reasons to support the argument (Ennis & Weir, 1985) received high scores. For example, one student wrote, "Mr. R's forth [sic] point has no backing as is purely his own skewed viewpoint."

On the other hand, another student received a low score because he/she did not see these flaws; in fact, he/she agreed with the statement made and gave no reason(s) as to why. He/She wrote, "Number four is a brief and concise reason why parking on the street is not good for the city."

Recognition of Assumption

Recognition of Assumption was based on scores on Paragraph 6. Participants who highlighted that an assumption was made based on a flawed experiment due to its short duration, poor sampling, and/or lack of controls (Ennis & Weir, 1985), received high

scores. One student noted two assumptions: the conclusion drawn from unreliable data, and the conjecture that everyone knows how many accidents occurred on a certain street in a given year. He/She wrote:

> In paragraph six Robert uses limited data from an experiment to buck his point. The data collected is [sic] not even close enough to making a decision/conclusion. Also the last sentence of this paragraph uses an assumption which makes his attitude relaxed and unreliable.

Participants who agreed with the assumption made and/or failed to see the problems with the experiment received low scores. For example, one student wrote, "Prohibiting parking in that time period would be fine. Accidents were probably decreased, but again if a driver is paying attention to what their [sic] doing, there wouldn't be an issue."

Interpretation

Interpretation was based on scores on Paragraph 7. Participants who were able to detect illegitimate use of the term "safe" to win an argument (Ennis & Weir, 1985) received high scores. One student wrote:

> ...Mr. Raywift gives his definition of what 'safe' is. However, we all know that streets, cars, cities are not safe. Eliminating parked cars during a time when everyone is sleeping will not remove "even the slightest possible chance for an accident".

Students who accepted the author's definition of "safe" and his emphasis on making conditions completely safe, received low scores. For example, one student wrote, "I agree with what was stated fully. Safe is that there is no chance in being harmed."

Summary

In this quantitative study, students came from three programs offered at two different schools. One school was chosen because it offered the IB program, the other because it offered the option to take AP classes. All participants were enrolled in the third quarter of their last year of high school. A total of 60 students in the study: 20 IB students, 20 AP students, and 20 NON students.

This study took place during one school day at each school. During one class period, students took a test of CT. Data were collected by the researcher immediately following completion of the test. Tests were scored in accordance with a rubric provided by the authors of the test. The researcher scored 100 percent of the tests. For inter-rater reliability, she trained a colleague to score tests using the same rubric. Her colleague also scored 100 percent of the tests. The rate of inter-rater reliability was 86 percent.

A series of chi-square analyses indicated that there were no significant differences between the IB, AP, and NON students with regard to gender, native language, parental education level, and family income. A series of one-way ANOVAs indicated that the groups were similar in terms of SAT scores and the number of years of attendance at the current school. GPA differed across groups; however, cumulative GPA and scores on tests of CT have been shown to be unrelated (Conger & Mezza, 1996; Cox, 2008; Leaver-Dunn, Harrelson, Martin & Wyatt, 2002; Scott et al., 1998; Williams et al., 2003). Information concerning these demographics was used to equalize the demographic characteristics of the three groups to reduce the possibility that differences in test scores across groups could be explained by demographic differences (Gay et al., 2008).

A series of one-way ANOVAs was conducted with academic programs as independent variables; and total scores, subscale scores, and use of emotive language scores on the EW as the dependant variables.

CHAPTER FOUR: RESULTS

Overview

Chapter Four describes the research questions and data analysis procedures used in the study. It outlines group scores in total points, paragraph scores, subscale scores, and scores in Use of Emotive Language to Persuade earned on *The Ennis-Weir Critical Thinking Essay Test* (EW). This chapter concludes with a summary of results as described in answers to the study's two guiding questions.

Research Questions

This study sought to address two questions:

1. What differences in critical thinking skills, if any, exist between students enrolled in academic programs that purport to foster critical thinking, namely the IB and AP programs, and students in a control group?
2. What differences in critical thinking skills, if any, exist between the critical thinking skills of students enrolled in the IB or AP program, both of which purport to foster critical thinking?

Method of Data Analysis

First, an analysis of variance (ANOVA) was conducted to examine differences between mean group scores in total points on the EW.

Total Points

Significant differences were found across groups in Total Points on the EW. The means (with standard deviations in parentheses) of Total Points are presented in Table 4.1.

The highest score possible on the EW was 29 points. Total Points differed across programs, $F(2, 57) = 8.493$, $MSE = 46.962$, $p = .001$. A follow-up Tukey HSD revealed that IB and AP groups had significantly higher total scores than the NON group. The IB and AP groups did not differ.

Table 4.1: Total Points scored on the EW by program.

	IB	AP	NON
Total Score	16.85 (7.8)	14.25 (5.7)	8.15 (6.9)

Individual Paragraph Scores

A second series of ANOVAs were conducted to examine differences on individual paragraph scores on the EW.

As described below, significant differences were found across groups in Paragraphs 4 (Inference), 6 (Recognition of Assumption), and 7 (Interpretation). The groups means (with standard deviations in parentheses) on Individual Paragraphs are indicated in Table 4.2. The asterisk indicates a significant difference across groups.

Paragraphs 1-8 on the EW were worth three points each. Paragraph 9 was worth five points because it summarized the preceding eight paragraphs and therefore included more information.

Paragraph 1 included the misuse of an analogy (Ennis & Weir, 1985). Paragraph 1 (with Paragraphs 2 and 5) tested the subscale of Evaluation of Argument, the ability to distinguish between strong and weak arguments (Watson & Glaser, 2006). Scores on Paragraph 1 did not differ across groups, $F(2, 57) = 1.951$, $MSE = 2.571$, $p = .151$.

Table 4.2: Scores on Individual Paragraphs on the EW by program.

	IB	AP	NON
Paragraph 1	1.6 (1.635)	1.05 (1.761)	.6 (1.391)
Paragraph 2	1.65 (1.755)	1.4 (1.634)	.50 (1.82)
Paragraph 3	1.6(1.5)	.95 (1.538)	.95 (1.731)
Paragraph 4*	2.05 (1.468)	2.0 (1.521)	.60 (1.635)
Paragraph 5	1.90 (1.586)	1.95 (1.394)	1.65 (1.424)
Paragraph 6*	2.20 (1.239)	2.05 (1.605)	.800 (1.765)
Paragraph 7*	2.55 (.998)	2.30 (1.260)	.950 (1.731)
Paragraph 8	1.50 (1.732)	1.55 (1.571)	1.05 (1.503)
Paragraph 9	1.80 (1.472)	1.20 (.833)	1.0 (.973)

Paragraph 2 included an irrelevant point to make an argument (Ennis & Weir, 1985). Paragraph 2 (with Paragraphs 1 and 5) tested the subscale of Evaluation of Argument. Scores on Paragraph 2 did not differ across groups, $F(2, 57) = 2.40$, $MSE = 3.024$, $p = .098$.

Paragraph 3 included a sound argument (Ennis & Weir, 1985). Paragraph 3 (with Paragraph 8) tested the subscale of Deduction, the ability to extract information in order

to form a sound conclusion (Ennis et al., 2005). Scores on Paragraph 3 did not differ across groups $F(2, 57) = 1.110$, $MSE = 2.539$, $p = .337$.

Paragraph 4* included circularity and/or poor reasoning (Ennis & Weir, 1985). Paragraph 4 tested the subscale of Inference, the ability to draw a conclusion based on observation or supposition of fact (Watson & Glaser, 2006). Paragraph 4 differed across groups, $F(2, 57) = 5.690$, $MSE = 2.382$, $p = .006$. A follow-up Tukey HSD revealed that IB and AP groups has significantly higher scores than the NON group. The IB and AP groups did not differ.

Paragraph 5 included a conclusion drawn without considering other possibilities (Ennis & Weir, 1985). Paragraph 5 (with Paragraphs 1 and 2) tested the subscale of Evaluation of Argument. Scores on Paragraph 5 did not differ across groups. $F(2, 57) = .239$, $MSE = 2.163$, $p = .788$.

Paragraph 6* included a faulty experiment (Ennis & Weir, 1985). Paragraph 6 tested the subscale of Recognition of Assumption, the ability to recognize when a conclusion has been drawn based on inadequate information (Ennis et al., 2005). Scores on Paragraph 6 differed across groups, $F(2, 57) = 4.904$, $MSE = 2.410$, $p = .011$. A follow-up Tukey HSD revealed that IB and AP groups had significantly higher scores than the NON group. IB and AP groups did not differ.

Paragraph 7* included an argument won by an inaccurate definition of a word (Ennis & Weir, 1985). Paragraph 7 tested the subscale of Interpretation, the ability to judge whether logical conclusions can be drawn with certainty based on given information (Watson & Glaser, 2006). Scores on Paragraph 7 differed across groups, F(2,

57) = 7.960, MSE = 1.861, p = .001. A follow-up Tukey HSD revealed that IB and AP groups had significantly higher scores than the NON group. IB and AP did not differ.

Paragraph 8 included a sound argument (Ennis & Weir, 1985). Paragraph 8 (with Paragraph 3) tested the subscale of Deduction. Scores on Paragraph 8 did not differ across groups, F(2, 57) = .588, MSE = 2.577, p = .558.

Paragraph 9 was a summary of the preceding eight paragraphs and did not correspond to a specific subscale. Scores on Paragraph 9 did not differ across groups F(2, 57) = 2.729, MSE = 1.270, p = .074.

Subscale Scores

A third series of ANOVAs were conducted to examine differences in subscale scores on the EW. Subscale scores were determined by the paragraph(s), worth three points each, that corresponded to each area of CT. Evaluation of Argument was comprised of Paragraphs 1, 2, and 5, and was therefore worth nine points. Deduction was comprised of Paragraphs 3 and 8, and was therefore worth six points. Inference was comprised of Paragraph 4, and was therefore worth three points. Recognition of Assumption was comprised of Paragraph 6, and was therefore worth three points. Interpretation was comprised of Paragraph 7 and was therefore worth three points. Paragraph 9 was not included in the subscale results because it was a summary and did not correspond to a particular subscale.

Significant differences were found across groups in Inference, Recognition of Assumption, and Interpretation. The group means (with standard deviations in

parentheses) on subscales are presented in Table 4.3. The asterisk indicates a significant difference across groups.

Scores on Evaluation of Argument did not differ across groups, $F(2, 57) = 2.681$, $MSE = 11.247$, $p = .077$.

Scores on Deduction did not differ across groups, $F(2, 57) = .964$, $MSE = 6.295$, $p = .388$.

Scores on Inference* differed across groups, $F(2, 57) = 5.690$, $MSE = 2.382$, $p = .006$. A follow-up Tukey HSD revealed that IB and AP groups has significantly higher scores than the NON group. The IB and AP groups did not differ.

Scores on Recognition of Assumption* differed across groups, $F(2, 57) = 4.904$ $MSE = 2.410$, $p = .011$. A follow-up Tukey HSD revealed that IB and AP groups had significantly higher scores than the NON group. IB and AP groups did not differ.

Scores on Interpretation* differed across groups, $F(2, 57) = 7.960$, $MSE = 1.861$, $p = .001$. A follow-up Tukey HSD revealed that IB and AP groups had significantly higher scores than the NON group. IB and AP did not differ.

Table 4.3: Subscale Scores on the EW by Program.

	IB	AP	NON
Evaluation of Argument	5.15 (3.745)	4.40 (3.03)	2.75 (3.242)
Deduction	3.10 (2.573)	2.50 (2.724)	2.0 (2.20)
Inference*	2.05 (1.468)	2.0 (1.521)	.60 (1.635)
Recognition of Assumption*	2.20 (1.239)	1.05 (1.605)	.80 (1.765)
Interpretation*	2.55 (.988)	2.30 (1.260)	.95 (1.731)

Use of Emotive Language to Persuade

An additional finding that emerged from the results of this study was the recognition of the author's use of emotive language to persuade his audience. In keeping with the EW scoring rubric (Appendix F), one point was awarded to participants who identified this language in statements such as "Any intelligent citizen should agree" or "These people don't know what safe really means" (Ennis & Weir, 1985).

The final ANOVA was conducted to examine differences between scores on Use of Emotive Language to Persuade. Significant differences were found across groups. The group means (with standard deviations in parentheses) on Use of Emotive Language to Persuade are presented in Table 4.4.

Table 4.4: Use of Emotive Language to Persuade Scores on the EW by Program

	IB	AP	NON
Use of Emotive Language to Persuade	.500 (.512)	.150 (.366)	.100 (.307)

Scores of Use of Emotive Language to Persuade differed across groups $F(2,57) = 5.791$, $MSE = .164$, $p = .005$. A follow-up Tukey HSD revealed that the IB group had significantly higher scores than the AP group and the NON group. AP and NON did not differ.

Research Questions

To address these questions and summarize results, information is presented in terms of total points scored on the EW, paragraph/subscale scores, and scores concerning use of emotive language to persuade readers.

Question One

What differences in critical thinking skills, if any, exist between students enrolled in academic programs that purport to foster critical thinking, namely the IB and AP programs, and students in a control group?

In terms of total points scored on the EW, both IB and AP scores were significantly higher than NON. There were significant differences found in three of the five subscales. In areas of Inference, Recognition of Assumption, and Interpretation, IB and AP scores were significantly higher than NON. In the subscales of Evaluation of Argument and Deduction, IB, AP, and NON scores did not differ.

In the category of Use of Emotive Language to Persuade, only IB scores were significantly higher than NON scores; AP and NON did not differ.

Question Two

What differences in critical thinking skills, if any, exist between students enrolled in the IB or AP program, both of which purport to foster critical thinking?

In terms of total points scored on the EW, IB and AP scores did not differ. There were no significant differences between IB and AP scores in the five: Evaluation of Argument, Deduction, Inference, Recognition of Assumption, and Interpretation. The

only significant difference between IB and AP scores was found in the category of Use of Emotive Language to Persuade; IB scores were significantly higher than AP scores in this area.

CHAPTER FIVE: DISCUSSION

Overview

The final chapter discusses the results of the study as they relate to the guiding questions. Following a discussion of the findings for each question, this chapter reviews limitations, addresses implications for practice, and makes recommendations for further research.

Introduction

The primary goal of this study was to determine if there were any differences in critical thinking (CT) abilities between students enrolled in either of two credit-based transitional programs that purport to develop CT skills, the International Baccalaureate (IB) program or the Advanced Placement (AP) program, and students in a control group. Credit-based transitional programs are programs designed to offer Grade 12 students opportunities to take college-level coursework and earn college credit while still in high school (Bailey & Karp, 2003).

CT is a pressing topic because of the widespread agreement that it should be considered one of the most important goals in education (Bernard, Zhang, Abrami, Sicoly, Borokhovski & Surkes, 2008; Ennis, 1993; Erwin & Sebrell, 2003; Giancarlo, Blohm & Urdan, 2004; Noddings, 2006, 2008; Siegel, 1986; ten Dam & Volmann, 2004; van Gelder, 2005; Willingham, 2007). CT skills are not only central to academic performance,

but are needed in social, professional, and interpersonal situations (Ku, 2009). CT skills equip students with the ability to think through complex issues in an ever-changing world (Paul, 1993).

Despite the agreed upon importance of CT, many schools fall short in teaching these valuable skills (Paul, 1993; Reed, 1998). This may be due in part to lack of a collective definition of CT (Paul, 1993); instruction that focuses more on memorization and recall than on CT (Gallagher, 1998); and/or assessment that tests the ability to recall facts, to the detriment of CT (Noddings, 2004).

The purpose of this study was to assess students' application of CT skills in analyzing and responding to arguments, in an effort to discern whether this ability could be attributed to the academic program in which a student was enrolled. Students enrolled in either the IB or AP program were recruited because these programs purport to develop students' CT skills (The AP Advantage, n.d.; Barbour & Streb, 2008; IBO, 2008). The control group consisted of students who were recruited because they were not in either program (NON), but were enrolled in challenging classes at the same school as the AP students.

All students took the same test, *The Ennis-Weir Critical Thinking Essay Test* (EW), which required reading a fictitious letter to the editor concerning overnight parking, and responding to arguments presented in the letter. The EW was selected for this study because it was found to be the instrument most aligned with the tenets of CT: it assesses students' abilities to problem-solve in a real world context; requires consideration of possibilities and use of evidence to support an argument; and elicits individual

interpretations and responses (Werner, 1991). CT tests in essay format allow for more complex thought processes than do multiple-choice tests (Brady, 2008; Brunt, 2005).

To examine the differences between the CT skills of students enrolled in different academic programs, the researcher attempted to answer the following guiding questions:

1. What differences in critical thinking skills, if any, exist between students enrolled in credit-based transitional programs that purport to foster critical thinking, namely the IB and AP programs, and students in a control group?
2. What differences in the critical thinking skills, if any, exist between students enrolled in either the IB or the AP program, both of which purport to develop critical thinking?

Discussion of the Research Questions

Question One

This question addressed the differences between scores of students in credit-based transitional programs (i.e., IB or AP students) versus scores of students in a control group (i.e., NON students). Results were based on group scores and are presented in order of total points on the EW; scores per subscale (i.e., Evaluation of Argument; Deduction; Inference; Recognition of Assumption; and Interpretation); and scores in the area of emotive language.

The scores of students in credit-based transitional programs were higher than those of students in a control group in terms of overall scores on the EW. This might

suggest that in general, the two credit-based transitional programs whose intent it is to develop CT are working.

Significantly higher scores among IB and AP students might be attributed to the EW's likeness to some of the assignments or forms of assessment that might be found in IB and AP classes. For example, AP teachers encourage the application of CT skills by having students find editorials they disagree with and write their own letters in response to them (Barbour & Streb, 2008). To do so, students must apply CT skills to analyze text and guide their own writing as they consider their audience, present their arguments clearly, and provide supporting evidence (Barbour & Streb, 2008).

Likewise, IB teachers incorporate past exams, some of which include analyzing and responding to articles or editorials, as instructional tools because they are open-ended essay questions that develop students' writing skills; teach them to support statements with evidence; and encourage them to make connections between different disciplines (Rothman, 2002).

These findings suggest that AP and IB students might be prepared for a test like the EW because they may have received instruction geared toward developing CT skills and been assessed using open-ended exercises where application of CT skills is necessary. However, because this study assessed students' skills but did not investigate ways in which students were taught, it is not certain whether or not all or some of the students in the study received this form of instruction.

No significant differences were found in the subscales of Evaluation of Argument and Deduction.

There were no significant differences between the scores of students in credit-based transitional programs and those of students in a control group in the subscale of Evaluation of Argument (Paragraphs 1, 2, and 5). It is surmised that scores on this subscale did not differ because the arguments presented in these paragraphs are fairly obvious, and students merely had to judge them strong or weak. Paragraphs 1 and 2 introduce the problem, and Paragraph 5 transitions the letter to paragraphs with more specific arguments (see Appendices E and F for test and scoring rubric).

Perhaps it is fairly evident in Paragraph 1 that cars parked in the street are not equivalent to garages. Had the author provided more reasons for this stance, it may have been more difficult to disagree with his claim; however, he did not. Most students were able to recognize the weakness of this argument.

Paragraph 2 follows closely the letter's introduction; it does not present a convincing argument that problems concerning afternoon rush hour traffic are related to problems concerning parking between 2 and 6 am. Most students were able to detect the irrelevance, and therefore weakness, of this argument.

In Paragraph 5, a sweeping generalization about eliminating accidents is made. Most students were able detect the weakness of this argument and generate possibilities outside of the narrow scope of it.

Likewise, there were no significant differences between the scores of students in credit-based transitional programs and those of students in a control group in the subscale of Deduction. Although Deduction is slightly more challenging than Evaluation of Argument in that it entails an additional step of extracting information from the argument

provided, it concerns reacting to given information, as opposed to something more complex, such as constructing a completely new argument.

Paragraphs 3 and 8 are sound arguments (see Appendices E and F for test and scoring rubric). Most students agreed with the author by deducing in Paragraph 1 that if there were fewer parked cars on the roads, that there would be more room for traffic. Secondly, in Paragraph 8, most students deduced that the information provided by two reliable sources supports the author's stance on prohibiting street parking from 2 to 6 am.

The fact that there were no significant differences between the scores of students enrolled in programs that aim to develop CT skills and those of students in a control group might be explained by some of the negative criticism these programs have received.

The IB and AP programs have been criticized on the grounds that in order to teach curriculum equal to college-level survey courses, these programs pressure teachers to cover a vast amount of material in a limited period of time (Callahan, 2003). To do so, some IB and AP teachers might rely on lectures rather than hands-on activities under the pressure to deliver a broad spectrum of material more efficiently (Callahan, 2003). These findings might suggest that despite their reputations for teaching CT (The A.P. Advantage, n.d.; Barbour & Streb, 2008; IBO, 2008), these programs may not actually do so, or may not do so effectively, given that lecture-based instruction has been determined a less effective instructional method of teaching CT than more interactive methods, such as cooperative or problem-based learning (Anderson, Mitchell & Osgood, 2005).

Again, because this study focused on a single assessment of CT rather than an in-depth look at CT instruction, it is not certain whether IB and AP teachers tried to cover

too much material in too short a time, or if their classes were lecture-based or not. The same is true for NON teachers and classes.

Significant differences were found in the subscales of Inference, Recognition of Assumption, and Interpretation.

The scores of students in credit-based transitional programs were significantly higher than those of students in a control group in the subscale of Inference (Paragraph 4).

Paragraph 4 contains only two sentences. It presents a limited argument made using circular language (see Appendices E and F). More IB and AP students than NON students pointed out these flawed aspects of the argument, mentioned that no reasons were offered to support the claim, and argued convincingly that the conclusion drawn could not have been made based on such limited information.

The scores of student in credit-based transitional programs were significantly higher than those of students in a control group in the subscale of Recognition of Assumption (Paragraph 6).

Paragraph 6 presents an experiment flawed by its short duration and problematic cause and effect reasoning (see Appendices E and F). More IB and AP students than NON students were able to articulate the flaws in the experiment due to lack of controls and inadequate sampling. They denied that a gross assumption (i.e., that it is wise to prohibit overnight parking on all streets of a town) could be made based on a poorly constructed experiment.

The scores of students in credit-based transitional programs were significantly higher than those of students in a control group in the subscale of Interpretation (Paragraph 7).

For Paragraph 7, more IB and AP students than NON students were able to perceive the author's attempt to win an argument by definition and in doing so, judge his argument as a false conclusion (see Appendices E and F).

The testing of Use of Emotive Language to Persuade was not originally part of this study. It was incorporated after the researcher had used the scoring rubric at length and saw results emerge in this area.

Emotive language was found in the Introduction ("There are a number of reasons why any intelligent citizens should agree."); Paragraph 5 ("All intelligent citizens would regard the near elimination of accidents in any period as highly desirable."); Paragraph 6 ("Everyone knows, of course, that there have been over four hundred accidents on Marquand during the past year."); and Paragraph 7 ("The opponents of my suggestions have said that conditions are safe enough now. These people don't know what "safe" really means.") Students were given one point for commenting on this persuasive technique (see Appendices E and F).

Only the scores of IB students were significantly higher than NON students' scores. More IB students than NON mentioned the author's use of emotive language and discussed its effect on readers. Skill in this area might indicate the ability to perceive a more subtle form of criticality; the aptitude to do so might intimate possession of a more refined type of CT skill.

Subscales where there were significant differences between scores of students in either credit-based transitional program, IB or AP, and students in a control group included Inference, Recognition of Assumption, and Interpretation. These differences

might be attributed to the course materials designed to develop CT that are made available to teachers in these two credit-based transitional programs.

The College Board AP website lists 212 publications, inclusive of all AP disciplines. Two CT strategies specific to AP include "AP-PARTS" (n.d.) and "CLUES" (Barbour & Streb, 2008), which give insight into how CT might be taught while guiding students in analyzing text.

AP-PARTS (n.d.) stands for Author, Point of view, Place and time, Audience, Reason, The main idea, and Significance. This acronym provides a sequence of important questions, such as, who created the source? What is his/her position? Where and when was the source created? For whom was the source intended? Why was the source produced at the time? What are its main points? Why is it important? As a strategy used in the classroom, AP-PARTS encourages students to analyze text critically.

The CLUES model is used primarily by American government teachers but may be adapted for other disciplines (Barbour & Streb, 2008). CLUES consists of specific prompts for CT: Consider the source. Lay out the argument and the underlying values and assumptions. Uncover the evidence. Evaluate the conclusion. Sort out the political implications (Barbour & Streb, 2008).

The instructional strategies of AP-PARTS and CLUES exemplify the Infusion Approach (Ennis, 1989) to teaching CT, an effective method of teaching CT (Barak, et al., 2007; Plath et al., 1999; Sokol, et al., 2008) where students are provided explicit strategies to apply CT skills. These strategies are employed by AP teachers in an effort to combat lower-level "descriptive thinking" through modeling steps of CT, rather than simply telling students to think critically (Barbour & Streb, 2008).

Either of these strategies, AP-PARTS or CLUES, or other strategies like them, would likely help students to dissect the arguments in the EW. As for the AP group in this study, it is not certain whether these students' teachers taught CT with AP-PARTS or CLUES, or materials similar to them, or whether these strategies may have affected their scores on the EW. Likewise, it is not certain whether NON students' teachers did or did not use such strategies or whether these strategies (or lack of) may have affected their scores on the EW.

In the IB program, Theory of Knowledge is a required philosophy class that aims to "provide coherence by exploring the nature of knowledge across disciplines, encouraging an appreciation of other cultural perspectives" (IBO, 2009). Theory of Knowledge involves instruction based on the fundamental questions: What do you know? How do you know what you know? What is language and how does it enable us to communicate what we know? (IBO, 2009; van de Lagemaat, 2006; Mathews & Hill, 2005).

Examples of resources for teachers of Theory of Knowledge include van de Lagemaat's (2006) *Theory of Knowledge for the IB Diploma,* which was designed to encourage CT across academic disciplines and assist students in formulating questions, using precise language, presenting ideas with supportive evidence, arguing reasonably, and making sound judgments.

Additionally, there is a *Theory of Knowledge Course Companion,* which is designed to stimulate discussion about learning and knowledge from multiple perspectives and across disciplines. It includes discussion prompts meant to encourage reflection, CT, awareness of how knowledge is constructed, and guidance for students to

consider the implications of knowledge in matters of global concern (Dombrosky, Rotenberg & Bick, 2007).

Like AP-PARTS and CLUES in the AP program, coursework in Theory of Knowledge would probably assist students in analyzing the arguments in the EW. In addition, its component of studying what language is and how it is used to communicate ideas might explain why IB students' scores were higher than NON students' scores in the area of Use of Emotive Language to Persuade.

It is certain that the IB students in this study took Theory of Knowledge classes because Theory of Knowledge is a required course in the IB program. It may be assumed that IB students studied different uses of language in Theory of Knowledge; however, because this study did not investigate particular curricula or methods of instruction, it can not be said with certainty that they did, or that the philosophical study of language may have positively affected their scores in this area of the EW. In comparison, the NON group did not take Theory of Knowledge, but this does not mean that they did not take other classes where different uses of language were studied, or that their study of language (or lack of) may have affected their scores on the EW.

Question Two

This question addressed the differences between scores of students in either of the two credit-based transitional programs included in the study (i.e., the IB and AP program). Results are based on group scores and are presented in order of total points on the EW; points per subscale (i.e., Evaluation of Argument; Deduction; Inference; Recognition of Assumption; and Interpretation); and points in the area of emotive language.

There were no significant differences between the scores of IB students and those of AP students in terms of total points on the EW. This might suggest that these credit-based transitional programs are equally effective in their theory and practice of developing students' general CT skills.

Likewise, there were no significant differences in the five subscales tested: Evaluation of Argument, Deduction, Inference, Recognition of Assumption, and Interpretation.

There were no significant differences between the scores of IB students and those of AP students in Evaluation of Argument (Paragraphs 1, 2, and 5). This might indicate that these programs teach the ability to articulate weaknesses of arguments equally well.

There were no significant differences between the scores of IB students and those of AP students in Deduction (Paragraphs 3 and 8). This might imply that the IB and AP programs are equally effective in teaching students to deduce the logic of arguments.

There were no significant differences between the scores of IB students and those of AP students in Inference (Paragraph 4). This shows that IB and AP students were equally adept at recognizing the limitations of the unsupported argument in this paragraph.

There were no significant differences between the scores of IB students and those of AP students in Recognition of Assumption (Paragraph 6). IB and AP students demonstrated equal proficiency in recognizing and communicating the limits of the experiment's controls and sampling.

There were no significant differences between the scores of IB students and those of AP students in Interpretation (Paragraph 7). An equal number of IB and AP students

determined that the author had won an argument by definition and made a false conclusion.

The fact that there were no significant differences between the scores of IB students and those of AP students in Total Points; Evaluation of Argument; Deduction; Inference; Recognition of Assumption; or Interpretation might suggest that the IB and AP programs are equally effective at teaching students to think critically. The efficacy of these programs might be attributed to the fact that they both aim to teach CT as evidenced by their missions and by strategies such as AP-PARTS and CLUES in the AP program and required courses like Theory of Knowledge in the IB program.

As mentioned previously, this study did not investigate specific teaching techniques in these programs; therefore, it is not certain that the teachers of the AP students in this study used AP-PARTS or CLUES, or strategies like them. The same is not true for the IB group. All IB students in this study had completed coursework in Theory of Knowledge, which aims to explore the essence of knowledge across academic domains and encourage consideration of different cultural perspectives (IBO, 2009), because it is a required component of the IB program; however, how these students were taught in Theory of Knowledge classes was not evaluated in this study. It is not certain that equality in scores on the EW could be attributed to particular forms of instruction or courses taught by teachers in either program.

The scores of IB students were significantly higher than those of AP students, as more IB than AP students mentioned the author's use of emotive language and discussed its effect on readers. Use of Emotive Language to Persuade was the only area of the test

where a difference was found between the scores of students in the two credit-based transitional programs.

Again, one might speculate that it is the required component of the Theory of Knowledge course in the IB program that might account for this difference in language skill because one of the aspects of Theory of Knowledge is the study of language and how it enables us to communicate ideas (van de Lagemaat, 2006; Mathews & Hill, 2005). Although it is certain that all IB students took Theory of Knowledge classes, it is not certain how the classes were taught, or if/how the Theory of Knowledge teachers used resources geared toward teaching uses of language in the context of CT. Likewise, it is certain that AP students did not take Theory of Knowledge because it is a course unique to the IB program, but this does not mean that AP students did not receive instruction on different uses of language.

Summary of Conclusions

As seen in discussions of some of this study's findings, there might appear to be a connection between academic programming and certain CT skills. Because there were no significant differences in scores across groups in Evaluation of Argument and Deduction, it is accepted that almost all students—IB, AP, and NON—were capable of applying these skills. On the other hand, there were distinct differences between the scores of IB and NON and AP and NON in the areas of Inference, Recognition of Assumption, and Interpretation. Students in credit-based transitional programs that purport to teach CT, had higher scores than students not in such programs, in these areas. IB students also had higher scores than NON in Use of Emotive Language to Persuade.

The only difference in scores between IB and AP students was found in Use of Emotive Language to Persuade. IB scores were higher than AP scores in this area. With the exception of Use of Emotive Language to Persuade, the IB and AP programs might teach CT equally well. This may be due to the intention to teach CT shared by both programs; the resources available to teachers in these programs; and/or program requirements that study CT and language.

Limitations

The results of this study pertain to the population described. Because the sample size was relatively small (N = 60), and one test was administered once at only two schools, the results will not generalize to a larger population of high school students or to a population of the same size without similar demographics. Replications with other populations would help strengthen the findings of this study.

Another limitation of this study is the timeframe in which it took place. One measurement during one 50-minute period provides a glimpse rather than in broad view of students' CT skills. However, the fact that there were significant differences between group means suggests that a more extensive study might allow wider exploration of how students develop and exercise CT skills. A more in-depth study might also include an experimental group taught by teachers trained in, for example, Richard Paul's Model for Critical Thinking and any or all of the following: administering pre- and post-tests to observe any changes in skill level; observing classes to examine specific CT strategies; reviewing curriculum to investigate CT resources available to students and teachers; and/or conducting interviews with students and teachers to determine if/how CT is

learned and taught. Due to limitations in this area, it was not certain whether it was the program or other factors that affected students' scores.

Another limitation resides in the participant groups. The researcher had access to a relatively modest number of IB students, of which 20 participated in the experiment. Because of the small size of the IB school (Site A), all students were enrolled in the IB program. Therefore, there were no NON students to recruit at Site A. Results were drawn based on IB students at Site A, AP students at Site B, and NON students at Site B. It would have been preferable to include NON students from both Sites A and B in order to make comparisons between these two groups. A comparison group from Site A may have shed more light on whether it was the program or the school that impacted students' CT abilities.

This study discusses ways in which the IB and AP programs foster CT skills (The AP Advantage, n.d.; Barbour & Streb, 2008; IBO, 2008); however, because this study did not focus on how participants were taught, it was not possible to address if/how their CT skills may have been developed.

Differences between the two schools in the study might be another limitation. Site A was a private, independent school, Site B a private, Catholic school. It is possible that because of these differences, instruction was different at each school. Site A focused on academic excellence, innovative thinking, and creative self-expression, Site B on academic excellence, moral leadership, and social responsibility in allegiance to the Catholic church. Again, because this study focused on one form of assessment, rather than on CT instruction, it is not possible to know if/how these differences played a role.

Another potential limitation was the difference between the length of the course of study in the different programs. The IB spans two years, Grades 11 and 12; AP courses normally take place during one school year, Grade 12. The longer duration of the IB programs may have favored the IB participants in this study; however, it is likely that AP students received more than one year of AP instruction in the sense that their coursework prior to Grade 12 probably focused on preparing them for AP classes.

Another limitation may have been in the researcher's former employment at Site A. Although measures were taken to ensure objectivity, such as scoring exams blindly, it is possible that her past teaching experience, knowledge of the IB program, and acquaintance with students at Site A, may have affected her role in the experiment.

A final limitation of this study was the testing instrument and the data available concerning it. While the EW was chosen from among several commercial tests of CT and deemed the most appropriate one for this study, it is somewhat outdated. It would be have been preferable to administer a more modern test, with subject matter of greater interest and concern to high school students in the twenty-first century than overnight parking. For this reason, some of the scores may not have accurately reflected students' CT skills because they may not have put forth as much effort as they would have had the subject been more important to them. In addition, there is little normative data on the EW. The results of this study show how IB, AP, and NON students faired relative to each other, but not in comparison to other populations of students. Lastly, it would have been helpful if the EW manual provided examples of students' work ranging between good, mediocre, and poor. This may have provided examiners with more objective criteria with which to assess exams.

Implications for Practice

The background of this study expounds on earlier research concerning whether CT can be taught, and if so, by what means. In addition, this experiment responded to a question in current research as to the effectiveness of credit-based transitional programs such as the IB and AP programs. It demonstrates that high school students are capable of thinking critically and applying certain skills in order to problem solve and propose solutions.

A result that emerged from this study was that despite similar demographics, participants in the IB and AP programs had significantly higher scores than students in the control group, not across the board but in certain areas of CT. These findings might suggest that to develop thinking skills beyond evaluating arguments and making deductions, teachers may need to be more explicit in their teaching of CT, namely the skills of Inference, Recognition of Assumption, Interpretation, and different uses of language.

Differences in overall scores despite the similarity of the groups in the experiment might suggest that the IB and AP programs develop CT skills not only in theory, but in practice. Both programs have resources and/or requirements designed to teach CT explicitly. Examples of these include AP-PARTS (an acronymic prompt for Author, Point of view, Place and time, Audience, Reason, and The main idea, and Significance) (AP-PARTS, n.d.) and CLUES (an acronymic prompt for Consider the source, Lay out the argument and the underlying values and assumptions, Uncover the evidence, Evaluate the conclusion, and Sort out the political implications) (Barbour & Streb, 2008) in the AP

program, and Theory of Knowledge (a philosophy course focused on studying knowledge and language) in the IB program. Symbolic of Ennis's (1989) Infusion Approach, wherein CT is taught explicitly, and which has been shown in several studies to be an effective method of teaching CT (Barak et al., 2007; Finlay & Faulkner, 2005; Lee, 2006; Plath et al., 1999; Sokol, et al., 2008), program components such as these may equip IB and AP students with tools in the form of thought processes and terminology that enhance CT capacities.

Recommendations for Further Research

The findings of this study indicate a need for further research in the following areas:

1. What might be learned from programs such as the IB and AP that can be applied to other programs so that all students are offered an education that teaches them to think critically?
2. What might be learned from long-term studies in which the relationship between students' CT skills and their future academic progress and professional endeavors are explored?
3. What information could derive from a qualitative study that includes observing classes; reviewing curricula; and/or conducting interviews with students and teachers to explore perceptions of CT and ways it might be taught and assessed?

Summary

This study measured various CT skills of students enrolled in different academic programs. It examined similarities between research on effective CT strategies and programming designed to teach CT. The findings of this study might suggest that students who receive instruction geared toward developing CT skills are more likely to exhibit higher CT skill levels than those who are not, in certain areas of CT.

REFERENCES

Anderson, W.L., Mitchell, S.M. & Osgood, M.P. (2005). Comparison of student performance in cooperative learning and traditional lecture-based biochemistry class. *Biochemistry and Molecular Biology Education 33*, 387-393.

The AP Advantage: Six Reasons to Consider AP (n.d.). Retrieved December 31, 2008 from http://www.scholarships.com/Advanced-Placement-AP-Program.aspx

AP-PARTS (n.d.). Retrieved January 2, 2009 from http://socialscientist.us/nphs/pdfs/apparts.pdf.

Bailey, T. & Karp, M. (2003). *Promoting college access and success: a review of credit-based transitional programs.* New York: Teachers College Columbia University, Community College Research Center.

Bailin, S. & Siegel, H. (2003). Critical thinking. In N. Blake, P. Smeyers, R. Smith, & P. Standish (Eds.), *The blackwell guide to the philosophy of education* (pp. 181-193). Oxford, England: Blackwell Publishing.

Barbour, C. & Streb, M. (2008.). Clues to thinking about politics. Retrieved December 30, 2008 from http://apcentral.collegeboard.com/apc/members/features/25896.html.

Barak, M., Ben-Chaim, D. & Zoller, U. (2007). Purposefully teaching for the promotion of higher-order thinking skills: a case for critical thinking. *Research in Science Education, 37*(4), 353-369.

Bernard, R., Zhang, D., Abrami, P., Sicoly, F., Borokhovski, E. & Surkes, M. (2008). Exploring the structure of the watson-glaser critical thinking appraisal: one scale or many subscales? *Thinking Skills and Creativity, 3*(1), 15-22.

Bickford, D. (2008). Using testimonial novels to think about social justice. *Education, Citizenship and Social Justice, 3*(2), 131-146.

Boud, D., Cohen, R. & Sampson, J. (1999). Peer learning and assessment. *Assessment in Higher Education, 24*(4), 413–426.

Brady, M. (2008). Cover the material—or teach students to think. *Educational Leadership, 65*(5), 64-67.

Brown, M. & Keeley, S. (1998). *Asking the right questions: a guide to critical thinking.* Upper Saddle River, NJ: Prentice-Hall Inc.

Brunt, B. (2005). Models, measurement, and strategies in developing critical thinking skills. *The Journal of Continuing Education in Nursing, 36*(6), 255-262.

Callahan, C. (2003). Advanced placement and international baccalaureate programs for talented students in american high schools: a focus on science and mathematics. The National Research Center for the Gifted and Talented. University of Connecticut

Campbell, D. & Stanley, J. (1963). *Experimental and quasi-experimental designs for research.* Boston: Houghton Mifflin.

Clark, D.R. (2009). Bloom's taxonomy of learning domains: the three types of learning. Retrieved August 26, 2008 from http://www.skagitwatershed.org/~donclark/hrd/bloom.html.

Conger, M. & Mezza, I. (1996). Fostering critical thinking in nursing students in the clinical setting. *Nurse Educator, 21*(3), 11-15.

Cox, W.C. (2008). The health sciences reasoning test as a tool to aid in the admissions process. Paper presented at the annual meeting of the American Association of Colleges of Pharmacy. Retrieved March 4, 2009, from http://www.allacademic.com/meta/p261433_index.html

Dewey, J. (1909). *How we think.* Boston: DC Heath & Co.

Dolhenty, J. (2008). The importance of critical thinking skills in our world today. Retrieved February 19, 2009, from http://209.85.173.132/search?q=cache:eY94ruxzy3cJ:www.radicalacademy.com/criticalthinking.htm+critical+thinking+skills+importance&hl=en&ct=clnk&cd=3&gl=us

Dombrosky, E., Rotenberg, L. & Bick, M. (2007). *Theory of Knowledge Course Companion.* New York: Oxford University Press.

Dong, Y. (2006). Learning to think in english. *Educational Leadership, 64*(2), 22-26.

El Hassan, K. & Madhum, G. (2007). Validating the waston-glaser critical thinking appraisal. *Higher Education, 54*(3), 361-383.

Ennis, R. (1999). An annotated list of critical thinking tests. Retrieved August 28, 2008 from http://www.criticalthinking.net/CTTestList1199.html.

Ennis, R. (1989). Critical thinking and subject specificity: clarification and needed research. *Educational Researcher, 18*(3), 4-10.

Ennis, R. (1993). Critical thinking assessment. *Theory into Practice, 32*(3), 179-186.

Ennis, R. & Weir, E. (1985). *The ennis-weir critical thinking essay test: test manual.* Pacific Grove, CA: Midwest Publications.

Ennis, R., Millman, J. & Tomko, T. (2005). *Cornell critical thinking tests: administration manual (5th edition revised).* Seaside, CA: The Critical Thinking Company.

Erwin, T. & Sebrell, K. (2003). Assessment of critical thinking: ETS's tasks in critical thinking. *The Journal of General Education, 52*(1), 50-70.

Finlay, S-J. & Faulkner, G. (2005). Tête à tête: reading groups and peer learning. *Active Learning in Higher Education, 6*(1), 32-45.

Gallagher, S. (1998). The road to critical thinking: the perry scheme and meaningful differentiation. *NASSP Bulletin, 82*(595) 12-20.

Gay, L., Mills, G. & Airasian, P. (2008). *Educational research: competencies for analysis and applications.* Upper Saddle River, NJ: Prentice Hall.

Giancarlo, C., Blohm, S. & Urdan, T. (2004). Assessing secondary students' disposition toward critical thinking: development of the california measure of mental motivation. *Educational and Psychological Measurement 64*(2), 347-364.

Glaser, E. (1941). *An experiment in the development of critical thinking.* New York: Teacher's College, Columbia University.

Hannel, I. & Hannel, L. (1998). The seven steps to critical thinking: a practical application of critical thinking skills. *NASSP Bulletin, 82*(598), 87-93.

International Baccalaureate Organization (IBO). (2008). *IB learner profile booklet.* Cardiff, UK: Peterson House.

IBO (2009). *Diploma programme curriculum core requirements: theory of knowledge.* Retrieved January 23, 2009 from http://www.ibo.org/diploma/curriculum/core/knowledge/

Jackson, M., Ignatavicius, D. & Case, B. (2006). *Conversations in critical thinking and critical judgment.* Sudbury, MA: Jones & Bartlett Publishers.

Johnson, A. (2000). *Up and out: using creative and critical thinking skills to enhance learning.* Neeham, Heights, MA: Allyn & Bacon.

Ku, K. (2009). Assessing students' critical thinking performance: urging for measurements using multi-response format. *Thinking and Creativity 4*(1), 70-76.

Kuhn, D. (1999). Developmental model of critical thinking. *Educational Researcher, 28*(2), 16-25, 46.

Kurfiss, J. (1988). *Critical thinking: theory, research, practice, and possibilities.* San Francisco: Jossey-Bass.

Lampert, N. (2006). Enhancing critical thinking with aesthetic, critical, and creative inquiry. *Art Education, 59*(5), 46-50.

Leaver-Dunn, D., Harrelson, G., Martin, M. & Wyatt, T. (2002). Critical thinking predispositions among undergraduate athletic training students. *Journal of Athletic Training, 37*(4) 147-151

Lee, D. (2006). Academic freedom, critical thinking and teaching ethics. *Arts & Humanities in Higher Education, 5*(2), 199-208.

Marie, S. & Cooper, A. (2002). Classroom choices for enabling peer learning. Theory into Practice, 41(1), 53–57.

Mathews, J. & Hill, I. (2005). *Supertest: how the international baccalaureate program can strengthen our schools.* Open Court Publishing: Chicago.

Morris, L.L., Fitz-Gibbon, C.T. & Lindheim, E. (1987). *How to measure performance and use tests.* London: Sage Publications.

Noddings, N. (2004). High stakes testing: why? *Theory and Research in Education, 2*(3), 263-269.

Noddings, N. (2006). *Critical lessons: what our schools should teach.* New York: Cambridge University Press.

Noddings, N. (2008). All our students thinking. *Educational Leadership, 65*(5), 8-13.

Oermann, M. (1999). Developing and scoring essay tests. *Nurse Educator, 24*(2), 29-32.

Paul, R. (1993). *Critical thinking: what every person needs to know in a rapidly changing world.* Santa Rosa, CA: Foundation for Critical Thinking.

Paul, R. & Elder, L. (2004). *The thinker's guide to the nature and functions of critical and creative thinking.* Dillon Beach, CA: The Foundation for Critical Thinking.

Paul, R. & Scriven, M. (1996). *Defining critical thinking: a draft statement for the national council for excellence in critical thinking.* Retrieved August 26, 2008, from http://www.criticalthinking.org/university/univclass/Defining.html.

Paul, R., Fisher, A. & Notsch, G. (1993). *Workshop on critical thinking strategies.* Sonoma, CA: Foundation for Critical Thinking.

Pike, G. (1997). Assessment measures: the california critical thinking skills test. *Assessment Update, 9*(2), 10-11.

Plath, D., English, B., Connors, L. & Beveridge, A. (1999). Evaluating the outcomes of intensive critical thinking instruction for social work students. *Social Work Education, 18*(2), 207-217.

Reed, J. (1998). Effect of a model for critical thinking on student achievement in primary document analysis and interpretation, argumentative reasoning, critical thinking dispositions, and history content in a community college history course (doctoral dissertation, University of South Florida, 1998).

Rothman, R. (2002). A test worth teaching to: the ib's course guides and exams make a good marriage. *The American Educator, 26*(2), 28-36, 46.

Saskatchewan Education (n.d.). *Understanding the common essential learnings.* Retrieved August 23, 2008, from http://www.sasked.gov.sk.ca/docs/policy/cels/el4.html#e13e14.

Scott, J., Markert, R., Dunn M. (1998). Critical thinking: change during medical school and relationship to performance in clinical clerkships. *Medical Education, 32*(1), 14-18.

Siegel, H. (1986). Critical thinking as an intellectual right. *New Directions for Child Development, 33,* 39-49.

Sokol, A., Oget, D., Sonntag, M. & Khomenko, N. (2008). The development of inventive thinking skills in the upper secondary language classroom. *Thinking Skills and Creativity, 3*(1), 34-46.

ten Dam, G. & Volman, M. (2004). Critical thinking as a citizenship competence: teaching strategies. *Learning and Instruction, 14*(4), 359-379.

Tomey, A. (1999). Selected response test items. *Nurse Educator, 24*(5), 9-10, 13.

van Gelder, T. (2005). Teaching critical thinking: some lessons from cognitive science. *College Teaching, 45*(1), 1-6.

van de Lagemaat, R. (2006). Theory of knowledge for the ib diploma. New York: Cambridge University Press.

Wagner, T. & Harvey, R. (2006). Development of a new critical thinking test using item response theory. *Psychological Assessment, 18*(1), 100-105.

Watson, G. & Glaser, E. (2006). *The watson-glaser critical thinking appraisal.* Retrieved September 1, 2008, from http://pearsonassess.com/hai/Images/dotCom/HTC/WG/SampleQuestions.pdf.

Werner, P. (1991). The ennis-weir critical thinking essay test: an instrument for testing and teaching. *Journal of Reading, 34*(6), 494-495.

Williams, K.B., Glassnap, D.R., Tilliss, T.S., Osborn, J.B., Wilkins, K., Mitchell. S., Kershbaum, W. & Schmidt, C. (2003). Predictive validity of critical thinking skills for initial clinical dental hygiene performance. *Journal of Dental Education, 67*(11), p. 1180-1192.

Willingham, D. (2007). Critical thinking: why is it so hard to teach? *American Educator: The Quarterly Journal of the American Federation of Teachers, 31*(2), 8-19.

Wolfs, F. (1996). *Appendix E: introduction to the scientific method.* Laboratory Manual of the Department of Physics and Astronomy. University of Rochester, NY. Retrieved August 26, 2008, from http://teacher.pas.rochester.edu/phy_labs/appendixe/appendixe.html.

Yeh, S. (2001). Tests worth teaching to: constructing state-mandated tests that emphasize critical thinking. *Educational Researcher, 30*(9), 12-17.

Appendix A

Letter to Administrator

Jane Walther
1207 W. Fort St. #116 · Boise, ID 83702
(208) 333 8086 · jane_walther@yahoo.com

[Date]

Dear [Name of Administrator],

I am a doctoral student of education at Boise State University. I have designed a study to measure students' critical thinking skills. I am contacting you to describe my study and ask your permission to recruit students at [Name of School].

With your permission, I would like to administer a critical thinking test in March 2009. The Institutional Research Board at Boise State has approved my use of *The Ennis Weir Critical Thinking Essay Test*, which is a well-known test that measures the ability to make accurate inferences, recognize assumptions, deduce properly, interpret information, and evaluate arguments. I would need about 30 Grade 12 students from your school who would be willing to take the test. All consent forms and permission slips have been approved and I would be happy to share them with you prior to recruiting students. The Ennis-Weir test takes approximately 40 minutes to complete and can be done in the computer lab at school.

I will report only group scores, as opposed to individual scores. I can guarantee that all participants will be made anonymous and your school will be given a pseudonym to ensure its anonymity as well. Your school's results will be available at the end of the study if that is of interest to you. I believe only positive results could come of this study, as it may increase students' awareness of their critical thinking skills and may lend insight into the acquisition and assessment of these important skills.

I would welcome the opportunity to meet with you at your earliest convenience. My schedule is very flexible. If you are interested, please contact me at the e-mail address or phone number above.

Thank you,

Jane Walther

Appendix B

Recruitment Script

Hello everyone,

My name is Jane Walther and I'm a doctoral student at Boise State. I'm working on my dissertation, which is like writing a *really long* research paper. In order to write that paper, I had to come up with an experiment, something that's never been done before. I decided to create a study that involves measuring high school students' critical thinking skills. Mr./Ms. [*insert name of teacher*] was nice enough to allow me to come into his/her class and talk to you about this study and see if any of you might want to participate in it. So, feel free to stop me and ask me any questions as I go along…

The requirements are pretty straightforward. You would take a standardized critical thinking test next month. This test is an essay test. It would be taken in the computer lab during Mr.'s/Ms.'s [name of teacher] class. There is only 1 question and you have 40 minutes to answer it. The point is to find out how you make deductions, interpret information, and evaluate arguments. If you've ever taken debate, these terms probably sound familiar. But you don't have to know how to debate in order to take the test. I never took debate, but when I looked over the test, I figured I would do fine on it.

All right, so there are some important things to know if you're interested in the study:

- Everything—names, scores, data, whatever—is *all* confidential. Nobody will know your score but you. I won't even see your name because you'll be given a number that identifies you. I keep all scores under lock and key. At the end of the study when I write my dissertation, I'll be reporting on group results, not individual results.
- As for the results, they really have nothing to do with anything outside of this study. Whether you choose to participate or not, whether you feel like you got a high score or not, does not affect your grades, GPA, class standing, or anything else in any way. I would love it if you could participate, but no one will be upset with you if you don't.
- Also, if you sign up for the study and then, for whatever reason, feel like you'd really rather *not* be in it, you're free to withdraw at any time, even if you're in the middle taking the test. There's no pressure or consequences. Absolutely none.
- Lastly, if you do want to participate, I'll need to get your signature and your parents' signature for permission to take the test. I have these forms with me today and will pass them around. Remember, I can't let you take part in the study if I don't have a form signed by one of your parents or guardians.
- I know this is a lot if information to retain in just a short period, so I've attached a flier to the permission slips that has all the information you need to know about the study.

Thanks, everyone. As those papers are going around, please feel free to ask me any questions….

Appendix C

Assent Form

ASSENT TO PARTICIPATE IN RESEARCH
BOISE STATE UNIVERSITY

My name is Jane Walther and I am a student of education at Boise State University. I am conducting a research study to measure students' critical thinking skills. I am asking you to take part in this research study so I can learn more about how students develop critical thinking skills and different ways these skills can be assessed. The test will take less than 1 hour.

If you agree to be in this study, you will be asked to take an exam in February or March 2009 in the computer lab at your school. This test measures overall critical thinking skills based on the ability to make evaluate arguments and defend your point of view. There are no "trick" questions on the test. You do not have to answer any question you don't want to and you can stop participating at any time. No one will be able to know how you responded to the questions and your name will never be used. At the conclusion of the study, responses will be reported as group results only. Your individual results will be made available only to you.

Being in this study is entirely voluntary. Your decision whether or not to participate will not affect your grades, GPA, or class standing in any way.

Please talk about this study with your parents before you decide whether or not to participate. I will also ask your parents to give their permission for you to participate. Even if your parents say "yes" you can still decide not to participate. You may also change your mind before or during the test. No one will be upset with you if you don't want to participate or if you change your mind later and want to stop.

You may ask me any questions about this study. You can reach me at the office of my academic advisor, Dr. Keith Thiede, at (208) 426-1278.

By signing below, you are agreeing to participate with the understanding that your parents have given permission for you to take part in this project. You are participating in this study because you want to.

Please return this form along with your parent's/guardian's signed consent form to [name of teacher] by the end of the week. Please keep the additional copy for your files. Thank you.

_____
Print Name of Student

_____ _____
Signature of Student Date

_____
Print Name of Parent

_____ _____
Signature of Parent Date

THE BOISE STATE UNIVERSITY INSTITUTIONAL REVIEW BOARD APPROVED THIS PROJECT AUGUST 14, 2008 FOR THE PROTECTION OF HUMAN PARTICIPANTS IN RESEARCH. #108-08-066

Appendix D

Parent/Guardian Consent Form

BOISE STATE UNIVERSITY
Parent/Guardian Consent

Dear Parent or Guardian:

A. PURPOSE & BACKGROUND

My name is Jane Walther. I am a doctoral student of Dr. Keith Thiede from the College of Education at Boise State University. I am conducting a study to analyze high school students' critical thinking skills. I would like to ask for permission for your son/daughter to participate in my study.

B. PROCEDURES

Your son/daughter will take a test of critical thinking in January 2009. The test measures the ability to evaluate arguments and defend a point of view. There are no "trick" questions on this test. It takes less than 1 hour to complete and will be taken in the computer lab of your son's/daughter's school.

C. RISKS & DISCOMFORTS

All information will remain completely confidential. No student will be identified by name. You are able to remove your son/daughter from the study at any time and his/her grades, GPA, or class standing will not be affected in any way.

Confidentiality: My records will be handled as confidentially as possible. Only my advisor Dr. Keith Thiede and I will have access to test results. Results will be kept on a password protected computer in Dr. Thiede's office. When the research is complete, test results will remain protected for three years (per federal regulations) and then destroyed. No individual identities will be used in reports or publications that may result from this study.

D. BENEFITS

There will be no direct benefit to your child from participating in this study. However, the information gained from this research may help education professionals better understand how students' critical thinking skills can be taught and assessed.

E. COSTS

There will be no cost to you or your son/daughter as a result of taking part in this study.

F. PAYMENT

There will be no payment to you or your son/daughter as a result of his/her taking part in this study.

G. QUESTIONS

If you have any questions or concerns about participation in this study, please contact me or Dr. Keith Thiede at (208) 426-1278. If for some reason you do not wish to do this, you may contact the Institutional Review Board, which is concerned with the protection of volunteers in research projects. You may reach the board office between 8:00 am and 5:00 pm, Monday through Friday, by calling (208) 426 5401 or by writing: humansubjects@boisestate.edu or Institutional Review Board, Office of Research Compliance, Boise State University, 1910 University Dr., Boise, ID 83725-1138.

Should you or your son/daughter feel discomfort due to participation in this research, you should contact your health care provider.

H. CONSENT

PARTICIPATION IN RESEARCH IS VOLUNTARY. I understand that I can choose not to have my son/daughter participate in this study, or to withdraw my child from participating at any time. Declining participation will not interfere with my son's/daughter's grades, GPA, or class standing in any way.

I will discuss this research study with my son/daughter and explain the procedures that will take place.

I have a copy of this consent form to keep.

I give my consent to allow my son/daughter to participate:

Print Name of Parent/Guardian	Print Name of Son/Daughter
Signature of Parent/Guardian	Date

If your son/daughter will be participating in this study, please return this form *along with his/her assent form* to **[insert teacher's name]** by the end of the week. Thank you.

THE BOISE STATE UNIVERSITY INSTITUTIONAL REVIEW BOARD APPROVED THIS PROJECT AUGUST 14, 2008 FOR THE PROTECTION OF HUMAN PARTICIPANTS IN RESEARCH. #108-08-066

Appendix E

The Ennis-Weir Critical Thinking Essay Test

THE ENNIS-WEIR CRITICAL THINKING ESSAY TEST AN INSTRUMENT FOR TESTING AND TEACHING

DIRECTIONS

Read the letter to the editor of the Moorburg newspaper. Consider it paragraph by paragraph and as a total argument. Then write a letter to the editor in response to this one. For each paragraph in the letter you are about to read, write a paragraph in reply telling whether you believe the thinking good or bad. Also write a closing paragraph about the total argument. *Defend your judgments with reasons.*

Your answer should have nine numbered paragraphs. Numbers one through eight should give your reactions to paragraphs one through eight in the letter. Your paragraph number nine should give your overall evaluation of the letter considered as one total argument. Each paragraph, including the last, should contain your reason(s).

Spend about 10 minutes reading the letter and thinking about it. Then write for not more than 30 minutes (about three minutes for each of your short paragraphs). The maximum total time for the test is 40 minutes.

Do not forget to give your reasons in each paragraph. Please write clearly.

Sign your name to your letter. You are a local citizen, and this topic concerns you.

Remember, write *nine numbered paragraphs* and give reasons.

THE MOORBURG LETTER

230 Sycamore Street
Moorburg
April 10

Dear Editor:

Overnight parking on all streets in Moorburg should be eliminated. To achieve this goal, parking should be prohibited from 2 a.m. to 6 a.m. There are a number of reasons why any intelligent citizen should agree.

1. For one thing, to park overnight is to have a garage in the streets. Now it is illegal for anyone to have a garage in the city streets. Clearly, then, it should be against the law to park overnight in the streets.

2. Three important streets, Lincoln Avenue, Marquand Avenue, and West Main Street, are very narrow. With cars parked on the streets, there really isn't room for the heavy traffic that passes over them in the afternoon rush hour. When driving home in the afternoon after work, it takes me thirty-five minutes to make a trip that takes ten minutes during the uncrowded time. If there were no cars parked on the side of these streets, they could handle considerably more traffic.

3. Traffic on some streets is also bad in the morning when factory workers are on their way to the 6 a.m. shift. If there were no cars parked on these streets between 2 a.m. and 6 a.m., then there would be more room for this traffic.

4. Furthermore, there can be no doubt that, in general, overnight parking on the streets is undesirable. It is definitely bad and should be opposed.

5. If parking is prohibited from 2 a.m. to 6 a.m., then accidents between parked and moving vehicles will be nearly eliminated during this period. All intelligent citizens would regard the near elimination of accidents in any period as highly desirable. So, we should be in favor of prohibiting parking from 2 a.m. to 6 a.m.

6. Last month, the Chief of Police, Burgess Jones, ran an experiment which proves that parking should be prohibited from 2 a.m. to 6 a.m. On one of our busiest streets, Marquand Avenue, he placed experimental signs for one day. The signs prohibited parking from 2 a.m. to 6 a.m. During the four-hour period, there was **not one accident** on Marquand. Everyone knows, of course, that there have been over four hundred accidents on Marquand during the past year.

7. The opponents of my suggestions have said that conditions are safe enough now. These people don't know what "safe" really means. **Conditions are not safe if there's even the slightest possible chance for an accident.** That's what "safe" means. So, conditions are not safe the way they are now.

8. Finally, let me point out that the Director of the National Traffic Safety Council, Kenneth O. Taylor, has strongly recommended that overnight street parking be prevented on busy streets in cities the size of Moorburg. The National Association of Police Chiefs has made the same recommendation. Both suggest that prohibiting parking from 2 a.m. to 6 a.m. is the best way to prevent overnight parking.

I invite those who disagree, as well as those who agree with me, to react to my letter through the editor of this paper. Let's get this issue out in the open.

Sincerely,

Robert R. Raywift

Appendix F

Criteria and Scoring Sheet for the Ennis-Weir

ENNIS-WEIR CRITICAL THINKING ESSAY TEST

Student's Name _____ Total Score _____ Graded By _____

CRITERIA AND SCORING SHEET FOR THE ENNIS-WEIR

Robert H. Ennis and Eric Weir

See manual for interpretation and qualification of these criteria.

	Credit Given (maximum is 3 points per line except #9)
1. Recognition of misuse of analogy, and/or recognition of shift in meaning, and/or claim that incorrect definition has been stipulated.	
2. Recognition of irrelevance.	
3. Recognition that Paragraph Three is OK. (Neglecting the busy-streets limitation is not penalized here.)[A]	
4. Recognition of circularity, and/or recognition that no reason is offered. (Subtract one point from credit for interpreting "undesirable" as "not desired.")	
5. Recognition that there may be other ways of preventing accidents, and/or recognition that other things might be more desirable, and/or recognition that there probably isn't much traffic at that time, and/or recognition that other types of accidents are unaffected, and/or recognition that no evidence has been given that such accidents occur. (Other possibilities)	
6. Recognition of lack of controls, and/or inadequate sampling, and/or "only one case," and/or "post hoc fallacy." (Other possible explanation)	
7. Recognition of winning argument by definition, and/or recognition that a word has been made useless for empirical assertion, and/or claim that an incorrect definition has been asserted.	
8. Recognition that Paragraph Eight is OK. (Neglecting the busy-streets limitation is not penalized here.)[A] [C]	
9. One point for just condemning the overall argument; another point for reviewing or summarizing the responses to the other paragraphs in some reasonable way; two points for recognizing (anywhere) the error of concluding about all streets on the basis of reasons that relate only to busy streets;[A] and one point for noting (anywhere) that Raywift has attempted to push people around with his emotive language. Total possible: 5 points.	

A score of –1, 0, 1, 2, or 3 will be given for each of the first eight numbered paragraphs:[B]

–1 judges incorrectly (good or bad)[C]
–1 shows bad judgment in justifying
0 makes no response[D]
+1 judges correctly (good or bad), but does not justify[C]
+2 justifies semi-adequately
+3 justifies adequately

For Paragraph Nine, the range is –1 to +5.

[A]Do not penalize for failure to note busy-streets limitation in Paragraphs Three or Eight. If it is not noted at least somewhere, do not give the allotted 2 points in Paragraph Nine. If the limitation is noted in Paragraphs Three or Eight, credit should be granted at Paragraph Nine.

[B]These criteria are guidelines. The grader should use judgment in awarding points, subtracting for unspecified errors and adding for unspecified insights.

[C]Sometimes, something judged one way here will be judged another way by the test taker, and so well defended that a positive score (sometimes even +3) is warranted. The grader must use judgment. For example, a good argument could be mounted against Paragraph Eight.

[D]If the examinee makes a response, but the argument of the paragraph is not judged either good or bad and no reasons are given, count it as "no response."

Appendix G

Demographic Questionnaire

_____

(Please print your initials and last 4 digits of your phone number)

Demographic Questionnaire

For this research project, I am requesting demographic information. Due to the make-up of Idaho's population, the combined answers to these questions may make an individual person identifiable. I will make every effort to protect participants' confidentiality. However, if you are uncomfortable answering any of these questions, you may leave them blank.

Gender: M F **Current Grade:** 9 10 11 12

Name of your school: _____

How many years have you attended this school (including this year)? _____

What school did you attend prior to this one? _____

Academic Program enrollment:
(Please mark the box and fill in the blanks of the one that applies to you.)

☐ **Advanced Placement (AP)**

How many AP courses are you currently enrolled in? _____

☐ **International Baccalaureate (IB)**

How many IB courses are you currently enrolled in? _____

☐ **Other**

Name of Program: _____

How many of this program's courses are you currently enrolled in? _____

☐ **I am not currently enrolled in a specific program.**

Your native language:

Of your parent(s), what is the highest level of education :
(Please circle answer/fill in blanks.)

high school undergraduate (BA, BS, etc.) master's (MA, MBA, etc.)

doctorate (MD, PhD, etc.) professional degree: _____ (e.g., lawyer, dentist)

(over)

Your Zip Code: **Your Grade Point Average (GPA):**

Your Scores on: **PSAT** _____ **SAT** _____ **ACT** _____

Estimated Family Income:
(please circle)

Below $50,000

50,000 – $100,000

101,000 – $200,000

Above $200,000

I don't know