MEASURING EXPERIMENTAL DESIGN ABILITY: A TEST TO PROBE CRITICAL THINKING

Jennifer Lynn Sieberg

A Thesis

Submitted to the Graduate College of Bowling Green
State University in partial fulfillment of
the requirements for the degree of

MASTER OF SCIENCE

August 2008

Committee:

Karen Sirum, Advisor

Daniel Pavuk

Paul Morris

ABSTRACT

Critical thinking is something of value for all. The educational system has been putting an emphasis on students' development of critical thinking skills. Critical thinking is a self-regulating, intellectual process of purposefully analyzing and evaluating all available information in order to formulate a well reasoned conclusion. Scientific thinking is critical thinking that involves the use of the scientific method, a method used to validate scientific knowledge, in order to study or investigate nature or the universe (Schafersman 1994). Science educators can facilitate the development of scientific thinking skills through scientific inquiry. What better way to experience scientific inquiry than through the design and implementation of experiments? "Developing the ability to design an experiment is critical to understanding of the scientific process and in promoting critical thinking skills. This skill can be developed if students are allowed to work like scientists" (Garcia 1999). The purpose of this research is to design, implement and determine the effectiveness of our test for experimental design ability, a form of scientific thinking, in an introductory biology course, through the use of the Experimental Design Ability Test (EDAT). We chose to assess experimental design ability due to the scientific and critical thinking skills that are involved in the design process and because it is applicable to students' everyday lives. Our findings indicate that the EDAT is sensitive to improvements in experimental design ability, as students exposed to student-designed laboratories made significant gains in their experimental design abilities, whereas those who were not did not yield gains.

To my family who
have inspired and supported me all the way.
Thank you.

ACKNOWLEDGEMENTS

I am forever grateful for my dedicated mentor Dr. Karen Sirum. Her unique approach to guiding me through has helped to open my eyes to one of the most important goals of education and life, thinking critically. Thank you to Dr. Dan Pavuk who has provided me with advice from the simple details about the Masters degree in general to the challenging task of statistical analysis, thank you for your time and effort. I would like to thank Dr. Paul Morris for the time that he has put into serving on my committee and for the insightful conversations. Dr. Dan Weigmann was also most helpful to me with my statistical analysis, thank you for your patience. I wish to thank all of my peers for their support and words of advice through this experience. I must also thank my family who has been a blessing throughout this entire experience and all that have preceded it, thank you from the depths of my soul and all that I am. Thank you to God and all of my guardian angels that have been working very hard to keep me focused and driven in spite of obstacles. I cannot express my gratitude enough to everyone who helped me to achieve this goal. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Development of critical thinking is a primary objective of higher education (Paul, 2008). Critical thinking is highly valuable as well as empowering to those who utilize it. Several definitions of what critical thinking is exist. The Foundation for Critical Thinking (2008) includes elements such as conceptualizing and synthesizing information to direct ones actions in their critical thinking. Schafersman (1991) describes it as accurate thinking in search of applicable and reliable information about the world. Critical thinking has also been defined by the Delphi Report as, "purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based" (Facione and American Philosophical Association, 1990).

Critical thinking experts through the utilization of the Delphi method put the Delphi Report definition together. This method includes a large group of experts, in this case 46, who obtain, collaborate, and present their ideas and opinions pertaining to a particular topic, in this case critical thinking. These individuals spent almost two years to decide what core elements are central to critical thinking, what are the various critical thinking dimensions, how to assess critical thinking, and how to instruct critical thinking and it is this definition that is the most widely regarded and used.

Critical thinking is a self-directed mental ability that drives our ability to solve problems and make well-rounded decisions. Critical thinking has two components, the skill to analyze and evaluate information and the disposition towards using critical thinking (Facione, 1990). Critical thinkers are a benefit to themselves and society. They have a tendency to get higher grades and are better at reasoning through choices in their daily lives (Quitadamo and Kuntz, 2007, U.S. Department of Education, 1990). They have an advantage in the job market because employers are seeking individuals who routinely use those skills (Carnevale *et al.* 1990; Holmes and Clizbe, 1997; National Academy of Sciences, 2005). They look at situations from every angle and do their best to be fair in judgment. They are willing to discount what they previously thought to be true when confronted with conflicting evidence. Critical thinkers use their ability with intent and purpose to solve problems. They convey “honesty in facing one’s own biases, prejudices, stereotypes, egocentric or sociocentric tendencies” (Facione and American Philosophical Association, 1990). This good critical thinker is one who has the ability to think critically and who chooses to use it as part of their daily lives (Facione and American Philosophical Association, 1990).

Critical thinking skills are a set of cognitive abilities. To develop as thinkers students progress through stages of cognitive development. Formal operational is the final stage in Piaget’s (1983) theory of cognitive development in children. By the time students are entering college they should be in the final stage, formal operational. This stage begins around the age of 12 and extends into adulthood, and during this time period, people are developing the ability to think about abstract concepts. Skills such as logical thought, deductive reasoning, and systematic planning also come out during this stage (Piaget, 1983).

Perry (1970, 1981) has developed a model that describes the levels of cognitive development in college students. He explains that most college students start out as dualistic then progress to be multiplistic and may eventually transition into becoming relativistic. A dualistic student relies strongly on authority and believes that there are right and wrong answers to every question that is obtainable from those in authority. A multiplistic student realizes that there is more than one perspective to a problem, but they feel that each perspective is just as good as the other. They end argumentation or avoid it by implying that everyone is entitled to their own opinion. A "relativistic student sees knowledge as relative to particular frames of reference. They show a capacity for detachment, they look for the "big picture," think about their own thinking, and evaluate their own ideas as well as those of others" (Perry, 1970, 1981). These students are often aware of different perspectives and realize that they can and should question authority (Perry, 1970, 1981). They are displaying their higher order thinking skills similar to the evaluation stage of cognitive learning as describe by Bloom (1956).

Bloom (1956) has developed a taxonomy of significant learning. It begins with the lowest level, knowledge, which represents simple recognition and recall of facts, and increases in complexity of mental abilities to the highest order identified as evaluation, the ability to judge the value of material for a purpose (Bloom, 1956). Blooms taxonomy includes 6 levels, each identified as follows (Bloom, 1956):

1. Knowledge: As remembering previously learned information
2. Comprehension: Grasping the meaning of informational materials
3. Application: The use of previously learned information in new and concrete situations to solve problems that have single or best answers.
4. Analysis: The breaking down of informational materials into their component parts, examining such information to develop divergent conclusions by

identifying motives or causes, making inferences, and/or finding evidence to support generalizations.

5. Synthesis: Creatively or divergently applying prior knowledge and skills to produce a new or original whole.
6. Evaluation: Judging the value of material based on personal values/opinions, resulting in an end product, with a given purpose, without real right or wrong answers.

Bloom (1956) stated that higher order thinking skills can be built on students' ability to identify concepts and analyze and integrate multiple concepts to solve problems. This is an ability that science courses can develop in their students through exercises that are meant to develop scientific thinking and reasoning.

National science organizations have expressed their support for new science education initiatives that develop critical thinking skills in students that will also promote scientific literacy (American Association for the Advancement of Science, 1989; National Research Council, 1995; National Science Foundation, 1996). Scientific literacy implies a person is able to evaluate the quality of scientific information and pose and evaluate arguments based on facts and apply this information properly (National Research Council, 1996). It follows therefore, that if someone were scientifically literate they would also have to be a critical thinker. The routine use of critical thinking skills in science builds on one's ability to reason scientifically and become scientifically literate.

> "Scientific reasoning is used to denote consistent, logical thought patterns which are employed during the process of scientific inquiry that enable individuals to propose relationships between observed phenomena; to design experiments which

test hypotheses concerning the proposed relationships; to determine all possible alternatives and outcomes; to consider probabilities of occurrences; to predict logical consequences; to weight evidence, or proof; and to use a number of instances to justify a particular conclusion" (Steussy, 1984).

Lawson (2004) argues, "scientific reasoning consists of an overall pattern of reasoning, which can be characterized as hypothetico-deductive, as well as several sub-patterns. Inhelder and Piaget (1958) referred to these sub-patterns as formal operational schemata…". Most students are arriving at universities in the formal operational stage of cognitive development and are capable of developing their scientific thinking skills, therefore, it is important that they are given ample opportunity to do so. The ability to think or reason scientifically is becoming increasingly important in today's society, as there are many decisions to be made that often involve a scientific element, anywhere from which of the latest medications to use, to how one's actions can impact the environment. "When one uses the methods and principles of scientific thinking in everyday life--such as when studying history or literature, investigating societies or governments, seeking solutions to problems of economics or philosophy, or just trying to answer personal questions about oneself or the meaning of existence--one is said to be practicing critical thinking" (Schafersman, 1994).

If the goal is critically thinking citizens what can done to help students think like scientists? According to Lawson, (93) "Thinking skills develop as a consequence of provoked encounters with situations in which students struggle to answer questions and reflect on those answers and on the methods of obtaining those answers." One method of facilitating this process is through the use of student-designed experiments that give students a “minds-on” experience. Dewey (1963) and Fischer (1980) both put emphasis on the importance of students’ having appropriate experiences in order to develop multifaceted thinking abilities. These experiences need to build upon one another throughout a students’ college education. It is in the hands of the instructor to ensure that they are aware of their students’ abilities in order to tailor the learning experience around what they think will result in the optimum learning of their students.

How can educators know whether their students are truly learning the skills they want them to if they do not directly assess those skills? Handelsman *et al.* (2006) describes scientific teaching as “Teaching science in a way that (1) represents the nature of science as a dynamic, investigative process based on evidence, (2) engages a diversity of people in a collaborative process and (3) has clear learning goals in mind, uses methods and instructional materials designed to improve student learning, and evaluates the methods iteratively.” In order to maximize the effectiveness of teaching pedagogies one should assess the attainment of learning by students. As an educator one can best benefit their students by first deciding what to teach based on what the learning outcomes are, then directly assess the learning outcomes, and modify teaching and assessment methods accordingly. Teaching in this manner utilizes the instructors’ critical thinking skills in effort to facilitate the development of this skill in students.

Lawson has developed an assessment for scientific reasoning, the Classroom Test of Scientific Reasoning. This test measures students’ scientific reasoning ability, such as combinatorial reasoning, probabilistic reasoning, proportional reasoning, and their ability to isolate control variables (Lawson, 1978). The test also has questions involving the conservation of weight and displaced volume (Lawson, 1978) (Appendix A). This test was established to have face validity, that is, experts in the field agreed 100% that the items on the test required formal-operational reasoning (Lawson, 1978). Convergent validity was established through correlating the classroom test total score and a summed bending rods and balance beam task that resulted in a correlation of 0.76 ($p<0.001$) (Lawson, 1978). Factorial validity was also established through a principal-component analysis that indicated there were three principal factors being measured, formal reasoning, concrete reasoning and formal-operational reasoning.

The Classroom Test of Scientific Reasoning has been used by many including Coletta and Phillips (2005) to investigate the relationship between normalized gains on the Force Concept Inventory (FCI) for physics and scientific reasoning ability. These researchers found a positive correlation between the two assessments ($r = 0.51$, $p<0.0001$) suggesting that teaching to help students with conceptual understanding in physics also helps student make gains on the Lawson test. Drawbacks to using the Lawson test are that it requires students to understand terms that may be unfamiliar in order for them to be able to answer the questions and it is a multiple-choice test. “It has been suggested that research shift toward more qualitative, open-ended approaches to assessment of individuals’ understanding (of any concept) be applied to the assessment of individuals’ nature of science conceptions” (Lederman, 1998). Multiple-choice questions provide unintended corrective feedback to the students. If they think they know the answer to the question, but the answer is not one of the choices, they must rethink their answer thus skewing the students’ initial response.

The Group Assessment of Logical Thinking test (GALT) (Roadrangka, 1982) is another assessment of logical thinking. Similar to Lawsons Test of Scientific Reasoning, this test measures proportional reasoning, controlling variables, probabilistic reasoning, combinational analysis, and correlational reasoning. The GALT is composed mainly of multiple choice questions along with some open-ended questions. The GALT was used to evaluate logical thinking in an inquiry based learning geology course versus a traditional geology course. Researchers found that students enrolled in the inquiry based learning course made a statistically significant 6.3% improvement in average GALT scores over one semester (McConnell *et al*. 2003) whereas the students enrolled in traditional courses did not improve. This test was also used by Guertin *et al.* (2005) to evaluate Just-In-Time Teaching pedagogies, which incorporate the use of online questions outside of the class to allow the instructor to clear up any misconceptions. This study found that the students who participated in Just-In-Time Teaching did not make gains in the GALT. This instrument is similar to Lawson's scientific reasoning test with drawbacks including multiple choice questions and the use of terminology that may be unfamiliar to students, which could lead them to answering the question incorrectly when they may really have the cognitive skills to answer the question correctly otherwise.

Another critical thinking assessment instrument is the California Critical Thinking Skills Test (CCTST) (Appendix B). This test measures students' critical thinking skills of analysis-interpretation, inference, and evaluation-explanation (Facione, 1990, 1998). The face validity, or agreement amongst experts, was established for the CCTST. Researchers also established validity of this assessment by conducting the test in experimental and control groups classes where the experimental groups are campus approved critical thinking courses and the control groups are introduction to philosophy courses. Both groups were assessed with the CCTST and results evaluated. This method is used to detect whether the instrument is sensitive enough to detect changes in critical thinking skills. Researchers found that the experimental groups made statistically significant growth in critical thinking skills while the control groups did not. Drawbacks with using this test include it requiring students to understand terms which some students may not be familiar with, the fact that it is a multiple-choice test, it requires 45 minutes to administer taking time from class, it is expensive to use, and you must send it out to be scored which takes time and also adds to the cost. Student buy-in is important to consider when choosing and using an assessment instrument effectively. If students do not care about the assessment, or feel that it is of importance to them, they may not give their best effort.

Given the limitations of the scientific thinking instrument available and the importance of assessment in scientific teaching we decided to try to design an assessment instrument that is inexpensive and easy to score, while effective at measuring scientific thinking. We decided to investigate the specific ability to design an experiment because of the scientific and critical thinking that are required in the design process. “Developing the ability to design an experiment is critical to understanding of the scientific process and in promoting critical thinking skills. This skill can be developed if students are allowed to “work like scientists” (Garcia, 1999).

We created our own Experimental Design Ability Test (EDAT) and implemented it in a nonmajor introductory biology course to find out what we could learn about this test’s sensitivity to detecting growth in experimental design ability. Evaluating students’ ability to design an experiment at the beginning and end of a course gives insight to the gains students have made in their ability to think scientifically and solve problems over the course. The EDAT requires that students explain how they would go about determining whether they would accept the claim about a product in an open-ended question format (Appendix C). This is a type of problem that people face in their daily lives, for example, determining what kind of data are required before one can accept a claim about an herbal supplement

Scientific thinking is used in the process of designing an experiment. One must analyze and evaluate the problem at hand, and to decide how they are going to set up an experiment in order to come as close as possible to solving the problem. Students need to understand the importance of controlling variables, larger sample sizes, and limitations to the generalization of their results. The EDAT does not require students to use direct terms such as independent or dependent variables. They first have to recognize that they can do an experiment to evaluate the validity of the specified claim. They then guide us through their thinking process in how they would design the particular experiment.

The EDAT is a straightforward, open-ended assessment of the gains students make in their experimental design ability from the beginning of a course to the end. It only requires 15-20 minutes to administer and upon completion 40 tests can be scored in one hour. Another benefit is that it is freely available at no financial cost to us or anyone else who would like to utilize it. The format of the EDAT also follows the shift toward open-ended assessment as described by Lederman (1998). This is important because it demands that students think through the process of designing an experiment in their own minds without being cued in on what the correct answer might be. Open-ended questions also reduce measurement error by eliminating random guessing that may occur on a multiple-choice test. This format gives insight into students' thought process instead of just the end result of their thinking. Using the EDAT will allow education researchers another way to investigate students' critical and scientific thinking.

To learn more about what can be revealed through the use of the EDAT, in the same population of students we used another critical thinking assessment instrument. While it is important to assess the development of critical thinking skills in students, it is also important to consider a method to assess the other component of critical thinking, the disposition to use critical thinking skills. When teaching critical thinking it is important to facilitate the development of critical thinking skills along with the disposition to use critical thinking skills. "…a person can master critical thinking skills without being the least bit disposed to use them" (Esterle, 1993). A good critical thinker is one who has critical thinking skills and chooses to use them (Facione and American Philosophical Association, 1990). This dispositional aspect of critical thinking is often overlooked in the classroom. However, science education goals include helping students become "citizen scientists", that is individuals who have and use their scientific thinking abilities. To investigate the relationship of the EDAT to critical thinking disposition we considered instruments designed to measure this dispositional aspect.

Various measures of students' attitudes towards science exist. The Aikenhead and Ryan (1992) Views on Science-Technology-Society (VOSTS) instrument is a survey of attitudes toward science (Appendix D). It investigates how "students view the social nature of science and how science is conducted" (Aikenhead and Ryan, 1992). It was validated through the process of development, which was based on student views (Aikenhead and Ryan, 1992). One drawback is that it is time consuming to administer as it consists of 114 multiple-choice questions. This also raised the concern that students may lose interest part way through the assessment. Again it is important to consider student buy-in when choosing an assessment tool. We are exploring students' critical thinking and therefore are focusing on students' attitudes toward using their critical thinking throughout their lives. Another assessment tool for attitudes toward science is the Views About Science Survey (VASS) (Halloun and Hestenes, 1998) (Appendix E). This surveys students' views about knowing and learning science as well as its relationship to students understanding of science (Halloun and Hestenes, 1998). The questions are not multiple choice, or open-ended; they are Contrasting Alternatives Design (CAD), which requires students to balance between two contrasting alternatives. This mirrors that of a likert-type scale (Halloun and Hestenes, 1998).

The VOSTS and the VASS are instruments to assess students' attitudes toward science (See Appendices D & E). These instruments question students about how they feel about science or different aspects of science. For our research purposes, we decided to look at the dispositional aspect of critical thinking because the information we obtain from this kind of instrument gives insight into students' behavior rather than how they feel about science (See Appendix F). The California Critical Thinking Disposition Inventory (CCTDI) is a pre/post test assessment instrument we decided to use to find out if students that make gains in their ability to design experiments also are more disposed to use critical thinking (Facione *et al.* 1992) (Appendix F). This survey provides scores in seven subscales: 1) truth-seeking 2) open-mindedness 3) analyticity 4) systematicity 5) critical thinking self-confidence 6) inquisitiveness 7) maturity. This test measures students' attitudes toward these seven dispositional aspects of critical thinking (Facione *et al.* 1992). The validity of this assessment was first established through researchers' conversations with individuals in the target population pertaining to their attitudes, expressions and beliefs toward critical thinking. Researchers then piloted the test to eliminate questions that failed to discriminate among test takers, items that responses inversely correlated with the overall test score, and those that added little or nothing by way of further refinement of overall score (Facione *et al.* 1992). Previous researchers, Phillips *et al.* (2004) have used this test to assess their pharmacy students and found significant gains in dispositions across a two and a half year investigation. Ultimately, the CCTDI was designed to investigate students' attitudes toward critical thinking in seven different aspects providing insight into whether students grow in these elements of critical thinking. We decided to use the CCTDI along with the EDAT to investigate the relationship between our EDAT and critical thinking dispositions in nonmajor introductory biology students.

Many times course objectives include the development of critical thinking skills without any mention of the development of the disposition to use those skills. Teaching someone a skill is most beneficial if they also have the desire to use it. "For liberal education, as well as for professional preparation at the collegiate level, educators must commit to sharpening students' cognitive skills and strengthening their disposition toward critical thinking" (Giancarlo and Facione, 2001). There was consensus in a national survey of employers, policy-makers, and educators that the dispositional aspect of critical thinking as well as the skill itself should be included as a necessary outcome of a college education (Jones *et al.* 1995). Many researchers think that in order to develop critical thinking in students the disposition to use those skills needs to be fostered as well (Facione and Facione, 1994; Paul, 1995; Bailin *et al.* 1999; Daly, 2001). The development of critical thinking skills does not lead to the development of critical thinking dispositions (Facione and Facione, 1994, Ricketts and Rudd, 2004, Foundation for Critical Thinking, 2008). Methods to facilitate the development of both skill and disposition are necessary in order to produce a good critical thinker. Therefore, it is useful to investigate the relationship between students' performance on the EDAT and on the CCTDI.

The EDAT may be an important new assessment instrument that can help facilitate the design and implementation of learning experiences that foster the development of both critical thinking skills and dispositions in college students. Assessment should not be used as an evaluative end to what students learn, rather as a tool to drive what and how students learn (Angelo and Cross, 1993; Center for Excellence in Teaching, 1999; McKeachie, 1999; Banicky and Foss, 2000; ETS, 2003; Middle States Commission on Higher Education, 2005). Assessment can inform educators about what students are having difficulty learning, and how educators may improve their approach to teaching certain lessons. Assessment can be equally beneficial for students and educators. Through this research we wish to determine whether our assessment methods effectively measure experimental design ability, and how these data might be used to meet the educational goal of producing critically thinking college graduates.

METHODS

The EDAT was used in sections of Introductory Biology 104 at Bowling Green State University from the spring semester of 2007 to the spring semester of 2008. Three different types of sections were involved: those that fully incorporated interactive engagement teaching strategies as well as student-designed labs (Experimental teaching groups 1-4), sections with a combination of traditional lecturing and active learning activities that fully incorporated student-designed labs (Experimental teaching group 5) (Table 1). The third type of sections consists of traditional lecture and traditional lab methods (Traditional teaching groups 1-6) (Table 1). The experimental groups are referred to as such because they are utilizing experimental teaching strategies.

The experimental groups 1-5 were challenged with a variety of interactive activities, discussions, as well as having the traditional lab component replaced by activities that involved student-designed experiments based on Handelsman *et al.* (2002) desk-top biology labs. These groups met 3 times a week; once a week for 2 hours and twice a week for 75-minutes. In groups 1-4 the instruction integrated lab group learning activities, discussions, and lecture, which was limited to approximately 15-minutes of lecturing. It is important to note that experimental group 5 fully incorporated the lab exercises while implementing some active learning strategies throughout the semester as a supplement to traditional lecturing methods. The traditional groups include students enrolled in the traditional sections of the introductory biology course. These students attended 50-minute lectures three times a week and participated in traditional descriptive labs once a week for a 2 hour time period.

Three different instructors taught these different groups (Table 1). One instructor for the experimental groups 1-4 taught lecture and lab. Another instructor taught the experimental group 5 lectures and a teaching assistant that was trained for teaching the student-designed laboratory taught the lab section of this group. The same instructor for experimental group 5 taught the traditional groups 1, 2, & 6. The third instructor taught the traditional groups 3-5 (Table 1). Note that the traditional group 3 was an honors section.

The lab activities that the students participate in for experimental groups 1-5 provide background information posing a problem to the students. Students are then asked to propose a hypothesis and design an experiment to test their hypothesis regarding the problem. Every student-designed experiment ends with a written lab report that is then thoroughly critiqued and returned to students with their final grade for the experiment. The lab report includes their hypothesis, methods, treatment group, control group and variables, results, discussion, and brief conclusions, and references. They are different from lab reports in traditional labs because the students design the methods section for each lab, designate what the treatment and control groups are, as well as the control variables. This gives students experience in designing experiments. These lab reports are then graded using a rubric that is also made available for students to use as a guide through the writing process. Students write 6 lab reports throughout the semester.

We analyzed the utility of the EDAT by administering it to nonmajor introductory biology students enrolled in both experimental teaching and traditional teaching groups at the beginning of the course and again at the end. It should be noted that the pretest question is different from the posttest EDAT question. The basic format of the EDAT does not change but the context of the question posed to the student can be varied, so that the prompt seems different to the students and requires thinking about a different situation and different details. (Appendix C).

The EDAT pretest was administered during the first week of the semester in each of the participating sections. Points were not awarded for participation in the pretest. Students are often eager to do their best at the beginning of the semester, which is why we do not award points for participation in the pretest. Pretest scores and the scoring rubric were not shared with students. Students were informed that we were gauging their abilities in biology to gain a better understanding of where they were and how we could help them be successful in this course.

The EDAT posttest was administered during the week prior to final exams in the experimental groups, and counted as a quiz grade, which is 5% of students' grade in the course. In the traditional groups the EDAT posttest was also administered during the week prior to final exams, however, the students in the traditional groups were informed that their score does not count toward their grade but they would earn bonus points for their effort and participation administration of the post EDAT in order to encourage them to put forth their best effort.

The EDAT was scored using a rubric that was adapted from existing experimental design rubrics for the purpose of simplification and clarification of the criteria for a good experimental design (University of Michigan-Dearborn 2002; Bizzell and Bizzell 2007; Science Olympiad 2005; Garcia 1999; Allen and Tanner 2006; Moskal 2000). The rubric score is based on 10 criteria for good experimental design (Appendix C) and student scores reflect the number of elements correctly included in each answer. The raters using the designed rubric scored the EDAT. Scores were tabulated independently then were compared for inter rater reliability. The inter rater reliability was determined to have a Pearson's Correlation Coefficient of 0.83 (Table 2). The reliability of this assessment is in the process of being established. A larger sample size will assist in the establishment of the reliability of this assessment. Its reliability thus far can be demonstrated by the fact that all of the sections that incorporated student-designed experiments made statistically significant gains. The validation of this assessment has been established thus far through scoring the EDAT with a scoring rubric that consists of a list of the elements that make a good experimental design (Appendix C), therefore establishing face or qualitative validation. Further validation is necessary through the use of other measures of experimental design, such as analysis of scientific thinking skills revealed in the lab reports. Statistical analysis of results was done so by utilizing Minitab 15 Statistical Software (2008).

The California Critical Thinking Disposition Inventory (CCTDI) (Facione and Facione 1992 and Insight Assessment 2006) was also a pre/post test assessment strategy to determine whether there is a relationship between CCTDI scores and EDAT scores. The CCTDI was administered during the first week of the semester in each of the participating sections. Students were not awarded any points for their participation. Students are given question and answer sheets and are told that the CCTDI is an instrument to look at their attitudes, beliefs, and opinions. The CCTDI was administered during the last week prior to final exams in all participating sections. Some traditional sections decided to give students extra credit in order to entice them to participate. The students' answer sheets are then sent to Insight Assessment in California, where they are scored, descriptive statistics are tabulated, and then sent back to us for evaluation. Minitab 15 Statistical Software (2008) was utilized for the statistical analysis of data obtained from the CCTDI. The correlation analysis between the EDAT and the CCTDI (Spearman's rank correlation) was done so by using STATISTICA (2008).

RESULTS

The EDAT was given in a pre and posttest format to students in nonmajor Introductory Biology 104 students. Only scores for subjects that participated in both the pre and posttest are reported and were used in statistical analysis. Various sections of the introductory biology course, including both experimental teaching groups and the traditional teaching groups, were assessed with the EDAT (See Table 1 for composition of various groups). The data did not fit a normal distribution, therefore I used a nonparametric test (One-sample Wilcoxon sign rank test) to analyze the data.

All of the sections with student-designed experiments made statistically significant gains in the EDAT ($p<0.001$), and the sections with traditional labs did not make gains (Table 3). The average EDAT pretest scores for the experimental and traditional groups were 3.6 (SD=1.8) and 3.3 (SD=1.9) respectively (Figure 1 & Table 4). The average EDAT posttest scores for the experimental and traditional groups were 6.6 (SD=1.35) and 4.0 (SD=1.5) respectively (Figure 1 & Table 4). This shows that for both groups, students' abilities are very similar at the beginning of the semester, however the gains that are being made are taking place in the experimental groups, indicating that something about the teaching methods utilized in this group facilitates the development of experimental design ability (Figure 1). Students that were exposed to laboratories that required them to design their own experiments and write detailed lab reports outperformed their peers that were exposed to traditional laboratory methods in their ability to design experiments as measured by the EDAT. This can be seen when looking at the distribution of the number of students with scores 1-10 on the EDAT pre and post for both experimental and traditional teaching groups (Figures 2 & 3). This average pretest score of a 3.6 or 3.3 doesn't indicate that students are receiving any 3 or 4 of the 10 criteria but typically the first 3 or 4. The rubric was designed incrementally, i.e., less experimental design skill necessary to obtain 1,2, and 3, more sophisticated experimental design skill required to get 4 and up.

Halpern (2001) found similar results when she assessed the effectiveness of critical thinking instruction and found that students that were taught with specific thinking instruction outperformed those that were not taught in this manner on standardized thinking skills tests. Although we were not assessing the effectiveness of the instructional method, the difference in instructional method used in the experimental and control groups gave us an opportunity to assess experimental design ability on classes that were and were not exposed to student-designed laboratory methods. Our finding is supportive of the effectiveness of our instrument as it is measuring experimental design ability, and many students that had an opportunity to use and develop this ability are the students that made gains in their EDAT scores.

The dispositional aspect of critical thinking is an integral part of critical thinking abilities (Facione and Facione, 1994; Paul, 1995; Bailin *et al.* 1999; Daly, 2001). We wanted to investigate the disposition to use critical thinking along with the development of experimental design abilities in order to determine whether there is some aspect of the introductory biology course that promotes the development of both ability and disposition that we could investigate further. We examined the changes in scores on the CCTDI from the pretest to the posttest for statistical significance (Table 5b). The samples for these tests include only the students who participated in both the pretest and the posttest.

Again the data did not fit a normal distribution so a nonparametric test (One-sample Wilcoxon sign rank test) was utilized.

Experimental group 3 made statistically significant gains in the truth-seeking subscale score, the mean gain was 2.57 points ($p=0.01$), experimental group 2 also made significant gains in this subscale ($p=0.06$), while all other groups did not show statistically significantly gains in truth-seeking (Table 5b). Experimental group 3 consisted of all first semester freshmen, while 57% of group 2 was freshmen (Table 1). Gains or losses in the truth-seeking subscale indicate that these students are either more or less disposed toward pursuing the truth (Insight Assessment, 2008). A truth seeker is more concerned with pursuing the truth than winning an argument and, they are eager to obtain the best information, even if it conflicts with their own beliefs (Insight Assessment, 2008).

Experimental group 1 made statistically significant gains, mean gain was 2.90 ($p=0.018$) in the confidence subscale, while all other groups did not make statistically significant gains in the confidence subscale score (Table 5b). This group was also made up of all first semester freshmen (Table 1). Gains in the confidence subscale indicate that students made gains in their trust toward their own reasoning (Facione and Facione, 1992). Therefore, 3 out of 5 experimental groups show statistically significant increases in their disposition toward critical thinking. One of the traditional teaching groups, group 4 made statistically significant gains in the maturity subscale, mean gain was 1.79 ($p=0.03$) throughout the semester (Table 5b). This group strengthened their positive disposition toward making reflective judgments. Out of the 5 experimental groups one significantly gained in truth-seeking when testing significance at $p<0.05$ and another (group 2) also significantly gained (mean gain was 2.15) in truth-seeking when testing significance at $p<0.10$. A third experimental group significantly gained in confidence ($p<0.05$). Only 1 out of the 6 traditional groups showed critical thinking disposition gains. This finding is encouraging because it indicates it is possible to provide an integrated learning environment that promotes the development of skills and dispositions that can lead to the production of good critical thinkers.

Critical thinking disposition changes may take longer than 16 weeks to detect as others, other researchers that have used the CCTDI looked at changes over 2-4 years. We thought we might be missing something when looking over the 16-week time period therefore, we looked at the trends of the CCTDI subscales scores (Figures 4-10). When evaluating the trends of the gains made in each subscale for each group the experimental groups are trending upward in their disposition more often than the traditional groups. Three out of five experimental groups have upward trends in the truth-seeking subscale score while only one out of six traditional groups have an up an upward trend in this subscale (Figure 4). This trend indicates that students are becoming more inclined to seek the truth in matters. College students generally enter in the dualistic stage as described by Perry (1970,1981). A dualistic student relies strongly on authority and believes that there are right and wrong answers to all questions that are obtainable from those in authority. Helping students to gain in their truth-seeking disposition early in their college education can stimulate their growth in the stages of cognitive development as they are in the beginning stages as described by Perry (1970, 1981).

Upward trends are not generally observed for either the experimental or traditional groups in systematicity, analyticity, and open-mindedness, inquisitiveness, or maturity (Figures 5,6,7,9 & 10). Three out of Five experimental groups have upward trends in confidence while only one out of six of the traditional groups show upward trends in confidence (Figure 8). The groups with inclining trends indicate that students in those groups are becoming more confident in their own reasoning. The upward trends in these dispositional aspects imply that if given more time the students could continue to become more positively disposed toward those aspects of critical thinking.

Facione and Facione (2001) found significant gains in dispositions across a four-year period. Specifically they found a significant improvement in truth-seeking by 2.8 points ($p<0.001$), confidence by 2.09 points ($p<0.001$), and total score by 7.43 points ($p<0.002$). Similar results were found in a study assessing pharmacy students' critical thinking (Phillips *et al.* 2004). This study found significant improvement in the total disposition score by 7 points ($p = 0.03$) across two and a half years, while there was no mention of significant gains in any of the subscales. While we found significant improvements in truth-seeking, confidence and maturity, further research is to reinforce the validity of our findings.

We wanted to determine if individual students who made gains in EDAT were the ones making gains in the CCTDI. Some research has found weak positive correlations between critical thinking skill and disposition whereas others have found that there is no correlation between skills and disposition (Giancarlo and Facione, 1994; Facione and Facione, 1997; Leaver-Dunn *et al.* 2002; Miller, 2003; Samawi, 2006). Giancarlo and Facione (1994) found a positive correlation ($r = 0.41$, $p<0.05$) between the CCTST, a critical thinking skills test, and the CCTDI in a study of 10th grade high school students.

Facione and Facione (1997) found a weak correlation ($r = 0.201$, $p<0.001$) between total scores in the CCTST and the CCTDI for nursing students at entry to their college programs. Similar correlations, never more than 0.194, were found when investigating a one-to-one relationship between specific critical thinking skills and dispositions (Facione and Facione, 1997).

We decided to investigate this relationship with our sample of students and the EDAT assessment instrument. The sample used in this correlation ($n = 67$) included experimental groups 3, 4, and 5. Two of these groups were all freshmen and one had a mixture of freshmen, sophomores and juniors. When using the Spearman's rank order correlation statistical analysis we found that overall the change in score for the EDAT did not statistically significantly correlate with the change in score for the CCTDI subscales when testing the significance at $p<0.05$. This may indicate that instructing for the development of critical thinking skills does not necessarily lead to the development of stronger critical thinking disposition in an individual student. It appears some other element influences critical thinking disposition. Similarly, students' change in score in the EDAT, a scientific thinking test, did not correlate with their change in score in the CCTDI.

We investigated differences between male and female gains in the EDAT and CCTDI to find if there are differences in gains among males and females. With the use of a two-sample t-test, no male-female differences were found in EDAT or CCTDI gains ($p=0.961$, $p=0.408$ respectively). This suggests that the teaching techniques were similarly beneficial or not for both male and female students and that the EDAT is not biased with regard to gender.

We also wanted to know if there is a statistically significant difference in CCTDI or EDAT prescores or gains among age (note: for our samples students ranged in age from18-25). Using a One-way ANOVA the data indicate that the mean scores of students ages 18-25 do not differ for either assessment. In other words, students did not come in with a higher score because of their age, and those students that did make gains did so regardless of their age.

One might think that students who have more college experience in general would perform better on the EDAT or the CCTDI. Students who have more college experience may have had more science courses or another course that promoted the development of their critical thinking. Using a One-way ANOVA we looked at the difference in prescores and gains made between freshman, sophomores, juniors or seniors on the EDAT and CCTDI. Results indicated that there was no difference. On average students, regardless of their year in college, are not entering introductory biology with the ability to score above 3.6 on the EDAT. Similarly students' average prescores were the same in the different CCTDI subscales regardless of their year in college. Although we did not find differences between prescores or gains for the undergraduate nonmajors enrolled in introductory biology for either EDAT or CCTDI based on gender, age, or year in college, differences may exist. Current work involves a larger sample of students that includes science majors.

It appears therefore, that the EDAT is sensitive to improvements made in experimental design ability, through the assessment of the experimental teaching groups which all made significant gains and the traditional teaching groups, which did not make gains. The gains made by individual students in EDAT score were found not to correlate with gains made in the CCTDI, which implies that the disposition to use critical thinking skills may not develop as a consequence or alongside the development of thinking skills in an individual student. This is consistent with a large body of research pertaining to skills and dispositions. However class averages that showed gains in EDAT scores also had average gains in CCTDI scores, indicating that teaching strategies involved in promoting experimental design ability can be integrated with strategies that promote critical thinking disposition.

The analysis of differences in prescores and gains in the EDAT and CCTDI among gender, age, and year in college indicate that there are no correlations, and that these variables did not influence prescores or gains. We found that the experimental groups made significant gains in the EDAT, and showed increasing trends in the CCTDI, but only experimental groups 1, 2 & 3 made statistically significant gains in the CCTDI subscales confidence ($p = 0.018$) and truth-seeking ($p=0.06$) and ($p = 0.048$) respectively. The traditional teaching groups did not show gains in the EDAT or increasing CCTDI trends except one group (Traditional group 4) made significant gains in their disposition toward maturity in making reflective judgments. It is possible that some of the group differences are due to differences in the groups in terms of ethnicity, incoming ACT or SAT scores, high school G.P.A., or previous science courses the students have had. So conclusions about the differences in outcomes among these groups are limited. However, this work demonstrates the utility of a new assessment tool, the EDAT, to this end.

DISCUSSION

The purpose of this study was to design, implement and determine the effectiveness of the EDAT, our test for experimental design ability, a form of scientific thinking, in an introductory biology course. Results indicate that students from the groups of the introductory biology course that incorporated student-designed experiments significantly made gains in their ability to design experiments compared to their peers enrolled in the groups that used traditional labs. All sections of the course started out the academic term with a similar low ability to design experiments. By the end of the term the students who were given practice in experimental design and critiqued in their experimental design made significant gains in their scores on the EDAT regardless of age, gender and year in college, or their initial disposition toward critical thinking as measured by the CCTDI. Current research is investigating whether these variables may have an influence when the sample size is larger and includes science majors.

The practice of formulating and conducting open-ended experiments throughout the semester appears to be contributing to students' ability to design experiments. The curiosity and investigation into the problem mirrors that of established scientists. Student-designed experiments allow students more freedom to explore and discover on their own in their own minds, hence helping the mind learn how to think through a challenge. Challenging someone to think for themselves by designing and executing their own experiment facilitates the process of critical thinking.

The analysis of changes in students critical thinking dispositions across the semester indicate that there is some aspect of the experimental groups that help students to significantly increase their truth-seeking and confidence dispositions and demonstrate overall increasing trends in disposition. Further research is necessary in order to determine what this may be, but the point remains that it is possible as well as necessary to integrate teaching strategies aimed at both goals. These data also show that students can make statistically significant gains in some dispositional elements of critical thinking within a short 16-week period, whereas previous researchers have found significant gains over years rather than a semester (Facione and Facione, 2001; Phillips *et al.* 2004).

A good question to ask is what kind of influence does the instructor have on the development of dispositions and critical thinking skills? Bellah and Dyer (2004) found that instructors are able to influence dispositions across a wide range of teaching strategies and styles. In our experiment we had a total of three instructors. One instructor taught only experimental 1-4 groups, another taught only traditional groups, and the other instructor taught both experimental 5 and traditional groups 1,2 & 5 (Table 1). For this instructor only the experimental teaching group that incorporated the student-designed labs and some interactive engagement learning did make statistically significant gains in the EDAT ($p<0.001$). This instructor's traditional group without the student-designed labs did not make gains in EDAT scores: average EDAT scores virtually remained the same (pre=3.66, post= 3.61) (Figure 1). This example points out that gains in the EDAT are not limited to one instructor. However, due to the small sample size at this point in the research project, it is not possible to say conclusively that this is not the case. Further work using the EDAT in many other courses with other instructors will help to clarify this issue.

Our assessment instrument indicates that allowing students to "work like scientists" has promoted the development of their experimental design ability, a form of critical thinking through designing experiments. The NRC (1999 & 2003) puts emphasis on learning through inquiry as it promotes learning, the development of skills involved in the practice of science, as well as enhancing students' retention of knowledge. Apedoe, *et al.* (2006) integrated inquiry-based learning into their geology course and found that students "increased dramatically" in their performance on inquiry-based exams. The professor of this course had previously used inquiry-based exams without implementing inquiry-based learning strategies and her students performed poorly on the exams. After the incorporation of inquiry-based learning strategies (activities and inquiry labs) she saw much improvement in their exam performance. While utilizing inquiry-based strategies in teaching are beneficial to students' development of critical thinking, a learning environment that promotes the development of both skill and disposition is necessary in order to produce critical thinking citizens who make use of their skill in their everyday lives.

Our assessment method was designed to measure students experimental design ability. It is to be expected that students enrolled in courses that give students training and practice in this skill would make larger gains in this ability. The students exposed to these student-designed laboratories did gain in their experimental design abilities, whereas those who were not did not yield gains. The EDAT offers an easy to administer and score measure of one aspect of scientific thinking, namely experimental design ability.

The ability to design experiments, if fostered, may help students become investigative on their own in other areas of their life. They may realize that there are many variables that have an influence on any particular phenomenon and it is important to be aware of each variable. This is similar to situations in life where decisions have to be made daily. More widespread use of the EDAT may help teachers get feedback on effective instructional strategies.

FUTURE DIRECTIONS

The dispositional aspect of critical thinking is just as important as the skills themselves. Figuring out a way to motivate students to care and want to use critical thinking is proving to be a challenging task; however, it is nothing that the scientists of today and tomorrow cannot overcome. It is important to use several different types of assessments that investigate the effectiveness of the individual types of activities that are used in a course in order to gain a deeper understanding of what these activities teach students. We have started by using the EDAT and the CCTDI. This investigation taught us that some of the strategies that we use are helping students gain in experimental design ability and critical thinking dispositions. Now we need to delve further to learn more about what is influencing gains in these abilities and dispositions.

This work has provided an assessment tool that has demonstrated how student-designed experiments help students to develop the ability to design experiments. Another important area for future research is to use another assessment of scientific thinking along side the EDAT. We know that scientific thinking is involved in the experimental design process, however it is important to use another assessment that directly measures scientific thinking to ensure the results from the EDAT are consistent with other measures. For example analysis of experimental design ability revealed in students' lab reports compared to EDAT scores could provide useful validation information. It is also necessary to use the EDAT with a greater sample of instructors and students to further establish validity of this instrument and increase the generalizability of our conclusions. In order to further develop students' science thinking skills one should also investigate data analysis skills since this is part of the accepted definition of scientific thinking and critical thinking (Schafersman, 1994; Facione and American Philosophical Association, 1990) and a valuable life skill as well. A more complete assessment of scientific thinking would be obtained by combing the EDAT with a data analysis test, part of the future research plans of this group.

Scientific teaching involves approaching the classroom as an experiment and asking if learning has taken place (Handelsman *et al.* 2004). The EDAT is an assessment instrument designed to measure experimental design ability. It can be used by teachers to provide feedback on how students are making gains in scientific thinking.

LITERATURE CITED

Aikenhead, G. S., & Ryan A. G. (1992). The Development of a New Instrument: Views on Science-Technology-Society (VOSTS). Science Education, 76, 477-491.

Allen, D., and Tanner, K. (2006). Rubrics: Tools for Making Learning Goals and Evaluation Criteria Explicit for Both Teachers and Learner. CBE-Life Sciences Education. Vol. 5, 197-203.

American Association for the Advancement of Science (1989). Science for All Americans. A Project 2061 Report on Literacy Goals in Science, Mathematics, and Technology, Washington, DC.

Angelo, Thomas A. and Cross, K. Patricia. (1993). Classroom Assessment Techniques: A Handbook for College Teachers (2nd ed.). San Francisco: Jossey-Bass.

Apedoe, X.S., Walker, S.E., Reeves, T.C. (2006). Integrating Inquiry-based Learning into Undergraduate Geology. Journal of Geoscience Education 54(3) 414-421.

Bailin S, Case R, Coombs JR, Daniels LB. (1999). Common Misconceptions of Critical Thinking. J Curriculum Stud. 31:269–283.

Bellah, K., Dyer, J.E. (2004). Influence of Leadership Style, Learning Style, and Leadership Adaptability on Critical Thinking Disposition. NACTA Journal. December.

Banicky, L.A. and Foss, H.K. (2000). Assessing Student Learning. Delaware Education Research and Development Center, University of Delaware.

Black, P. (2003). The Importance of Everyday Assessment. In: Everyday Assessment in the Science Classroom, ed. Atkin, J.M. and Coffey, J.E. Arlington: VA, NSTA Press. 1-11.

Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., and Krathwohl, D. R. (1956). A Taxonomy of Educational Objectives: Handbook 1, Cognitive Domain, New York: McKay.

Canas, A.J., Novak, J.D. (2006). Concept Maps: Theory, Methodology, Technology. Proceedings of the Second International Conference on Concept Mapping. San Jose, Costa Rica.

Carnevale, A. P, and American Society for Training Development (1990). Workplace Basics: The Essential Skills Employers Want, San Francisco, CA: Jossey-Bass.

Center for Excellence in Teaching. (1999). Teaching Nuggets. Los Angeles: University of Southern California.

Chaplin, S.B., Manske, J.M., and Cruise, J.L. (1997). Introducing Freshman to Investigative Research—a Course for Biology Majors. J. Coll. Sci. Teach. 27(5),347-351.

Coletta, V.P. and Phillips, J.A. (2005). Interpreting FCI Scores: Normalized Gain, Preinstruction Scores, and Scientific Reasoning Ability. American Jouranal of Physics. 73(12).

Cooper, J.L., and Robinson, P. (2000). Getting started: informal small-group strategies in large classes. In: Strategies for Energizing Classes: From Small Groups to Learning Communities: New Directions for Teaching and Learning , vol.81 , ed. J. MacGregor, J.L. Cooper, K.A. Smith, and P. Robinson. San Francisco: Jossey-Bass,17 -24.

D'Avanzo, C., and McNeal, A. (1997). Research for all Students: Structuring Investigation into First-year Courses. In: Student-Active Science: Models of Innovation in College Science Teaching , ed. A.P. McNeal and C. D'Avanzo. Philadelphia: Saunders College Publishing, 279-300.

Daly WM. (2001). The Development of an Alternative Method in the Assessment of Critical Thinking as an Outcome of Nursing Education. J Adv Nurs.36:120–130.

Davis, B.G. (1993). Tools for Teaching. San Francisco: Jossey-Bass.

Dewey, J. (1963). Experience and Education. New York: Collier Books. (Original work published in 1938).

Ennis, R. H. (1985). A Logical Basis for Measuring Critical Thinking Skills. Educ. Leadership 43(2), 44–48.

Esterle, J, Clurman, D. Conversations with Critical Thinkers. San Francisco: the Whitman Institute, 1993.

ETS (2003). Linking Classroom Assessment with Student Learning. Educational Testing Service. http://www.ETS.org

Facione, P. (1990, 1998). The California Critical Thinking Skills Test (CCTST): Forms A and B; and the CCTST Test manual. Millbrae, CA: California Academic Press.

Facione P. A., and American Philosophical Association (1990). Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. Research Findings and Recommendations, Millbrae, CA: Insight Assessment.

Facione, N., Facione, P. (1997). Critical Thinking Assessment in Nursing Education Programs: An Aggregate Data Analysis. Millbrae, CA: The California Academic Press.

Facione, P., Giancarlo, C., Facione, N., and Gainen, J. (1995). The disposition toward critical thinking, Journal of General Education, 44 (1), 1-25.

Facione, N., Facione, P., Giancarlo (Sánchez), C. (1994) Critical Thinking Disposition as a Measure of Competent Clinical Judgment: The Development of the California Critical Thinking Disposition Inventory. Journal of Nursing Education, 33(8), 345-350.L

Facione, P. and Facione, N. (1992). The California Critical Thinking Dispositions Inventory (CCTDI); and the CCTDI Test manual. Millbrae, CA: California Academic Press.

Facione, P. A., Facione, N. C., and Giancarlo, C. A. (1992). Test Manual: The California Critical Thinking Disposition Inventory, Millbrae, CA: Insight Assessment.

Foundation for Critical Thinking. (2008). (accessed 9 April 2008). http://www.criticalthinking.org

Fischer, K. W. (1980). A Theory of Cognitive Development: The Control and Construction of Hierarchies of Skills. Psychological Review, 87(6), 477-531.

Garcia, M, A. F., (1999). How to Design an Experiment. (accessed 17 April 2007). http://www.scienceteacherprogram.org/chemistry/Garcia99.html

Giancarlo, C. A., & Facione, P. A. (2001). A Look Across Four Years at the Disposition Toward Critical Thinking Among Undergraduate Students. The Journal of General Education, 50(1), 29-55.

Glasson, G.E., and McKenzie, W.L. (1998). Investigative Learning in Undergraduate Freshmen Biology Laboratories. J. Coll. Sci. Teach. 28,189 -193.

Grant, B.W., and Vatnick, I. (1998). A Multi-week Inquiry for an Undergraduate Introductory Laboratory. J. Coll. Sci. Teach. 28,109 -112.

Guertin, L.A., Zappe, S.E., & Kim, H. (2007). Just-in-Time Teaching (JiTT) Exercises to Engage Students in an Introductory-level Dinosaur Course. Journal of Science Education and Technology, 16: 507-514.

Halloun, I., and Hestenes, D. (1998). Interpreting VASS Dimensions and Profiles. Science and Education, 7(6), 553-577.

Handlesman, J., Houser, B., Kriegel, H. (2006). Biology Brought to Life: A Guidebook to Teaching Students to Think Like Scientists. McGraw-Hill.

Holmes, J., and Clizbe, E. (1997). Facing the 21st century. Bus. Educ. Forum. 52(1), 33–35.

Inhelder, B. and Piaget, J. (1958). The Growth of Logical Thinking from Childhood to Adolescence. New York: Basic Books.

Jones, E.A., Hoffman, S., Moore, L.M., Ratcliff, G., Tibbetts, S., & Click, B.L. (1995). National Assessment of College Student Learning: Identifying the College Graduate's Essential Skills in Writing, Speech and Listening, and Critical Thinking. Washington D.C.: National Center for Educational Statistics, US Department of Education, Office of Educational Research and Improvement. OERI Publication NCES 93–001.

Kitchen, E., Bell, J.D., Reeve, S., Sudweeks, R., and Bradshaw, W. (2003). Teaching cell Biology in the Large-enrollment Classroom: Methods to Promote Analytic Thinking and Assessment of their Effectiveness. Cell Biol. Educ. 2,180 -194.

Klionsky, D.J. (2004). Talking Biology: Learning Outside the Book—and the Lecture. Cell Biol. Educ. 3,204 -211.

Knight, J. K. and Wood, W.B. Teaching More by Lecturing Less
Cell Biol Educ 2005 4: 298-310.

Lawson, A.E. (2004). The Nature and Development of Scientific Reasoning a Sythetic View. International Journal of Science and Mathematics Education, 2(4), 307-338.

Lawson, A.E., and Johnson, M. A. (1998). What are the Relative Effects of Reasoning Ability and Prior Knowledge on Biology Achievement in Expository and Inquiry Classes? Journal of Research in Science Teaching, 35(1), 89-103.

Lawson, A.E. (1993). Deductive Reasoning, Brain Maturation, and Science Concept Acquisition: Are they Linked? Journal of Research in Science Teaching, 30, 1029–1052.

Lawson, A.E. (1978). The Development and Validation of a Classroom Test of Formal Reasoning. Journal of Research in Science Teaching, 15(1), 11-24.

Lederman, N. G., Wade, P. D. & Bell, R. L. (1998). Assessing the Nature of Science: What is the Nature of our Assessments? Science and Education, 7 (6), 595-615.

Leaver-Dunn, D., Harrelson, G.L., Martin, M., Wyatt, T. (2002). Critical Thinking Predisposition Among Undergraduate Athletic Training Students. Journal of Athletic Training. 37(4 suppl): S-147-S-151.

Lee, K. (2000) Childhood Cognitive Development The Essential Readings. Blackwell Publishers.

Lord, T.R. (1994). Using Constructivism to Enhance Student Learning in College Biology. J. Coll. Sci. Teach.23 , 346-348.

Lord, T.R. (2001). 101 Reasons for Using Cooperative Learning in Biology Teaching. Am. Biol. Teach.63 , 30-38.

Lynd-Balta, E. Using Literature and Innovative Assessments to Ignite Interest and Cultivate Critical Thinking Skills in an Undergraduate Neuroscience Course. CBE Life Sci Educ 5(2): 167-174 2006.

Mazur, E. (1996). Peer Instruction: A Users Manual. Upper Saddle River, NJ: Prentice-Hall, Inc.

McConnell, D.A., Steer, D.N., Owens, K.D. (2003). Assessment and Active Learning Strategies for Introductory Geology Courses. Journal of Geoscience Education. March.

McKeachie, W.J. (1999). Teaching Tips: Strategies, Research, and Theory for College and University Teachers. Boston: Houghton Mifflin.

Middendorf, J., and Kalish, A. (1996). The “Change-Up” in Lectures. National Teaching And Learning Forum. Vol. 5. No. 2. (accessed 14 April 2007). http://www.ntlf.com/html/pi/9601/v5n2.pdf

Middle States Commission on Higher Education. (2005). Assessing Student Learning and Institutional Effectiveness, Understanding Middle States Expectations. Philadelphia, PA.

Miller, D.R., (2003). Longitudinal Assessment of Critical Thinking in Pharmacy Students. American Journal of Pharmaceutical Education. 67(4) Article 120.

Minitab 15 Statistical Software. (2008). Minitab Incorporated. State College. PA.

Moore, R.W., & Sutman, F.X. (1970). The Development, Field Test, and Validation of an Inventory of Scientific Attitudes. Journal of Research in Science Teaching, 7, 85–94.

Moskal, B. M. (2000). Scoring Rubrics: What, When and How?. Practical Assessment, Research & Evaluation, 7(3). (accessed 14 April 2007). http://PAREonline.net/getvn.asp?v=7&n=3

National Academy of Sciences, National Academy of Engineering, Institute of Medicine (2005). Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future, Washington, DC: Committee on Prospering in the Global Economy of the 21st Century.

National Research Council (1995). National Science Education Standards, Washington, DC: National Academy Press. National Research Council (2003). Bio 2010, Transforming

National Research Council. (1996). National Science Education Standards. Washington, DC: National Research Council.

National Research Council (1999). How People Learn: Brain, Mind, Experience and School. Washington, DC: National Academic Press.

National Research Council (2003). BIO 2010, Transforming Undergraduate Education for Future Research Biologists, Washington, DC: National Academies Press.

National Science Foundation (1996). Shaping the Future: New Expectations for Undergraduate Education in Science, Mathematics, Engineering, and Technology, Washington, DC: Directorate for Education and Human Resources.

Paul, R. (2008). The National Council for Excellence in Critical Thinking. Foundation for Critical Thinking, Santa Rosa, CA. (accessed 8 March 2008). http://www.criticalthinking.org

Paul, R. (1995). How to Prepare Students For A Rapidly Changing World. Foundation for Critical Thinking. Santa Rosa, CA. (accessed 8 March 2008). http://www.criticalthinking.org

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). Knowing What Students Know: The Science and Design of Educational Assessment. Washington, DC: National Academy Press.

Perry, W. G. (1970). Forms of Intellectual and Ethical Development in the College Years, New York: Holt Rinehart & Winston.

Perry, W.G. (1981), 'Cognitive and ethical growth: the making of meaning', in A. W. Chickering (ed.), The Modern American College, San Francisco: Jossey Bass, pp. 76-116.

Phillips, C.R., Chesnut, R.J., Rospond, R.M. (2004). The California Critical Thinking Instruments for Benchmarking, Program Assessment, and Directing Curricular Change. American Journal of Pharmaceutical Education 68(4).

Piaget, J. (1983). Piaget's theory. In P. Mussen (ed.). Handbook of Child Psychology. 4th edition. Vol. 1. New York: Wiley.

Quitadamo, I.J. and Kurtz, M.J. (2007). Learning to Improve: Using Writing to Increase Critical Thinking Performance in General Education Biology Cell Biol. Educ. 6: 140-154.

Redish, E.F., Steinberg, R.N., Saul, J.M. (1998). Student Expectations in Introductory Physics. Am. J. Phys. 66, 212-224.

Ricketts, J.C. and Rudd R. (2004). The Relationship between Critical Thinking Dispostions and Critical Thinking Skills of Selected Youth Leaders in the National FFA Organization. Journal of Southern Agriculture Education Research 21. 54(1).

Roadrangka, V., Yeany, R., and Padilla, M., 1982, GALT, Group Test of Logical Thinking, Athens, Georgia: University of Georgia.

Samawi, R.N. (2006). The Effect of Concept Mapping and Critical Thinking Skills and Dipositions of Junior and Senior Baccalaureate Nursing Students. In Concept Maps: Theory, Methodology, Technology. Proceedings of the Second International Conference on Concept mapping, A. J. Canas and J. D. Novak (Eds.), San Jose, Costa Rica.

Schafersman, S. D. (1991). An Introduction to Critical Thinking. (accessed 9 April 2008). http://www.freeinquiry.com/critical-thinking.html

Schafersman, S.D. (1994). An Introduction to Science, Scientific Thinking and the Scientific Method. (accessed 9 April 2008). http://www.freeinquiry.com/intro-to-sci.html

Springer, L., Stanne, M. E., and Donovan, S. (1997). Measuring the Success of Small-Group Learning in college-Level SMET Teaching: A Meta-Analysis. National Institute for Science Education. (accessed 20 April 2007). http://www.wcer.wisc.edu/archive/cl1/CL/resource/scismet.htm

STATISTICA. (2008). StatSoft, Incorporated. Tulsa, OK.

Steer, D.A., Owens, K.D. (2003). Assessment and Active Learning Strategies for Introductory Geology Courses. Journal of Geoscience Education. McConnell, March 2003.

Stuessy, C. (1984). Path Analysis: A Model for the Development of Scientific Reasoning Abilities in Adolescents. Journal of Research in Science Teaching, 26(1), 41-53.

Sundberg, M.D., and Moncada, G.J. (1994). Creating Effective Investigative Laboratories for Undergraduates.BioScience 44,698 -704.

Tanner, K., and Allen, D. (2004). Approaches to Biology Teaching and Learning: From Assays to Assessments—On collecting Evidence in Science Teaching. Cell Biol. Educ. 3: 69-74.

U.S. Department of Education (1990). National Goals for Education, Washington, DC.

White, B., & Frederiksen, J. R. (1998). Inquiry, Modeling, and Metacognition: Making Science Accessible to All Students. Cognition and Instruction, 16(1), 3-118.

Wirth, K. R., and Perkins, D. (2007). Learning to Learn. (accessed 16 April 2007). http://www.macalester.edu/geology/wirth/CourseMaterials.html

TABLES

Table 1. Characteristics of Introductory Biology Course Sections

Section	Section Characteristics								
	Term	# of Ss**	N EDAT	N CCTDI	% Freshmen	% Female	@ Instructor	Lab Method*	Lecture Method*
Exp. 1	F '06	22	0	20	100%	36%	A	S-D	AL
Exp. 2	Sp '07	21	21	20	57%	57%	A	S-D	AL
Exp. 3	F '07	25	24	21	100%	80%	A	S-D	AL
Exp. 4	F '07	22	21	20	23%	64%	A	S-D	AL
Exp. 5	F '07	24	22	19	100%	58%	B	S-D	Trad. + AL
Trad. 1	Sp '07	ND	0	20	100%	70%	B	Trad.	Trad.
Trad. 2	Sp '07	ND	0	83	53%	77%	B	Trad.	Trad.
Trad. 3	F '07	ND	12	13	8%	62%	C	Trad.	Trad.
Trad. 4	F '07	ND	20	19	100%	56%	C	Trad.	Trad.
Trad. 5	F '07	ND	76	62	40%	66%	C	Trad.	Trad.
Trad. 6	Sp '08	ND	71	58	48%	52%	B	Trad.	Trad.

Note: Exp. = Experimental teaching groups. Trad. = Traditional teaching groups.
*Lab method is either S-D (Student-Designed) or Trad. (Traditional). Lecture method is AL (Active Learning), Trad. (Traditional) or both.
** # of Ss indicates number of students enrolled in course which varies from sample number. ND = not determined
- % Freshman & % Female are tabulated from the sample of students that participated in assessments.
@ Groups were taught by three different instructors indicated by A,B, or C.

Table 2. Determination of inter-rater reliability value for EDAT (Pearson's Coefficient)

EDAT Inter-rater Reliability		
$r = 0.835$	$p < 0.001$	
	M	*SD*
Rater 1	5.16	2.54
Rater 2	5.56	2.31

Table 3. Results from Wilcoxon Sign Rank Test: Determination of Significant Gains in EDAT for Each Participating Section.

Section	EDAT Change from Pre to Posttest			
	M	*SD*	*Median*	*P*
Exp. 2 n = 21	2.28	2.03	2.5	<0.001*
Exp. 3 n = 24	3.92	2.28	3.5	< 0.001*
Exp. 4 n = 21	3.81	2.16	4.0	< 0.001*
Exp. 5 n = 22	2.55	1.99	2.5	< 0.001*
Trad. 3 n = 12	0.17	2.25	0.00	0.894
Trad. 4 n = 20	-0.25	1.77	0.00	0.570
Trad. 5 n = 76	0.42	2.09	0.5	0.063
Trad. 6 n = 71	-0.06	1.87	0.00	0.768

* p<0.05

Table 4. The Means and Standard Deviations of Pre and Post test EDAT Scores

	Mean EDAT Scores			
	Pretest		Posttest	
Section	*M*	*SD*	*M*	*SD*
Exp. 2 n = 21	4.67	1.74	6.52	0.93
Exp. 3 n = 24	3.29	1.73	7.21	1.64
Exp. 4 n = 21	3.14	1.91	6.95	1.32
Exp. 5 n = 22	3.33	1.88	5.77	1.51
Trad. 3 n = 12	3.33	1.92	3.50	1.24
Trad. 4 n = 20	3.30	1.76	3.05	1.43
Trad. 5 n = 76	3.00	1.76	3.42	1.81
Trad. 6 n = 71	3.66	2.07	3.61	1.69

Table 5a. The Means and Standard Deviations of Pre and Post test CCTDI Scores

	Experimental 1 (*n* = 20)				Experimental 2 (*n* = 20)			
	Pretest		Posttest		Pretest		Posttest	
Measures	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
T	34.05	4.94	34.70	6.22	33.80	4.70	35.95	5.82
O	39.20	6.71	40.75	6.08	41.45	5.13	40.70	5.51
A	42.05	4.52	41.80	4.34	41.85	5.00	41.30	4.95
S	39.65	6.23	41.15	5.63	38.40	5.70	37.80	5.014
C	40.00	4.71	42.90	4.56	42.10	6.90	40.75	6.52
I	41.50	4.19	41.35	5.81	42.15	5.14	42.00	5.81
M	40.90	7.93	40.00	7.54	41.50	5.84	42.90	6.59

Note. T = Truth-seeking, O = Open-mindedness, A = Analyticity, S = Systematicity; C = CT Self-Confidence; I = Inquisitiveness; M = Maturity.

	Experimental 3 (*n* = 21)				Experimental 4 (*n* = 20)			
	Pretest		Posttest		Pretest		Posttest	
Measures	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
T	32.76	7.24	35.33	5.84	35.60	6.77	35.55	6.73
O	40.95	4.85	39.86	5.98	43.25	5.62	41.90	6.02
A	43.33	5.35	41.29	5.82	43.90	5.34	44.40	5.46
S	37.62	6.86	38.10	6.09	36.60	6.92	35.70	7.36
C	41.19	6.23	40.76	5.69	44.25	6.40	44.65	6.28
I	43.00	7.47	40.10	8.94	43.45	5.11	44.50	5.74
M	41.48	7.63	40.57	7.73	42.10	6.54	41.15	6.45

Note. T = Truth-seeking, O = Open-mindedness, A = Analyticity, S = Systematicity; C = CT Self-Confidence; I = Inquisitiveness; M = Maturity.

Table 5a Continued: *The Means and Standard Deviations for Pre and Post CCTDI Scores*

	Experimental 5 (*n* = 19)				Traditional 1 (*n* = 20)			
	Pretest		Posttest		Pretest		Posttest	
Measures	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
T	35.58	5.69	33.21	6.24	34.30	4.81	33.85	4.63
O	41.95	4.80	40.63	6.00	39.45	5.59	39.40	4.39
A	42.26	4.21	41.05	3.91	41.65	4.33	40.55	4.94
S	37.79	5.76	37.47	5.00	38.40	5.86	38.50	4.73
C	39.84	5.96	41.37	5.84	39.75	5.43	39.90	6.50
I	43.47	5.46	42.11	7.01	41.70	4.92	39.60	5.29
M	41.89	7.91	40.26	6.88	40.35	6.07	39.65	6.15

Note. T = Truth-seeking, O = Open-mindedness, A = Analyticity, S = Systematicity; C = CT Self-Confidence; I = Inquisitiveness; M = Maturity.

	Traditional 2 (*n* = 83)				Traditional 3 (*n* = 13)			
	Pretest		Posttest		Pretest		Posttest	
Measures	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
T	34.66	5.55	34.63	6.18	40.15	4.62	40.08	7.91
O	41.61	6.39	40.90	5.97	46.38	4.59	45015	4.54
A	42.11	5.52	41.59	5.87	45.38	5.38	46.23	6.10
S	39.36	7.19	39.00	6.79	40.46	7.04	40.04	7.37
C	42.25	6.21	42.16	6.15	43.00	4.88	44.85	7.36
I	41.40	7.35	40.69	6.36	44.85	4.86	45.31	7.31
M	41.88	7.83	40.72	7.92	45.62	5.97	44.69	4.71

Note. T = Truth-seeking, O = Open-mindedness, A = Analyticty S= Systematicity; C = CT Self-Confidence; I = Inquisitiveness; M = Maturity.

Table 5a Continued: *The Means and Standard Deviations for Pre and Post CCTDI Scores*

	Traditional 4 (*n* = 19)				Traditional 5 (*n* = 62)			
	Pretest		Posttest		Pretest		Posttest	
Measures	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
T	33.63	3.89	33.16	5.16	34.45	6.17	34.38	6.51
O	40.89	6.43	39.84	6.85	42.15	5.61	41.29	6.01
A	41.74	5.04	41.05	4.79	40.92	5.34	41.97	6.24
S	37.32	5.65	37.26	4.52	37.89	7.20	38.44	7.00
C	40.68	5.96	40.58	5.95	42.07	7.48	42.05	6.93
I	43.74	5.42	41.05	6.81	43.58	6.91	42.05	7.10
M	40.95	4.88	42.74	6.02	41.90	6.91	41.23	8.13

Note. T = Truth-seeking, O = Open-mindedness, A = Analyticity, S = Systematicity; C = CT Self-Confidence; I = Inquisitiveness; M = Maturity.

	Traditional 6 (*n* = 58)			
	Pretest		Posttest	
Measures	*M*	*SD*	*M*	*SD*
T	34.57	5.37	35.16	5.86
O	41.59	6.48	40.28	6.96
A	42.79	4.68	42.50	6.40
S	38.95	6.15	39.12	6.01
C	42.53	5.94	42.48	6.74
I	42.41	6.65	42.09	5.96
M	41.59	6.82	40.76	7.81

Note. T = Truth-seeking, O = Open-mindedness, A = Analyticity, S = Systematicity; C = CT Self-Confidence; I = Inquisitiveness; M = Maturity.

Table 5b. Results from Wilcoxon Sign Rank Test: Determination of Significant Gains within Each CCTDI Subscale for Each Participating Section

	Experimental 1 (*n* = 20) Change from Pre to Posttest			Experimental 2 (*n* = 20) Change from Pre to Posttest		
Measures	*M*	*SD*	*P*	*M*	*SD*	*P*
T	0.65	4.09	0.45	2.15	4.7	0.06**
O	1.55	5.19	0.25	-0.75	3.99	0.22
A	-0.25	5.39	0.57	-0.55	4.78	0.51
S	1.50	5.04	0.26	-0.60	4.06	0.48
C	2.90	4.84	0.018*	-1.35	5.91	0.48
I	-0.15	6.05	0.91	-0.15	6.03	0.91
M	-0.90	3.89	0.42	1.40	5.21	0.23

Note. T = Truth-seeking, O = Open-mindedness, A = Analyticity, S = Systematicity; C = CT Self-Confidence; I = Inquisitiveness; M = Maturity.

* $p<0.05$, ** $p<0.10$

	Experimental 3 (*n* = 21) Change from Pre to Posttest			Experimental 4 (*n* = 20) Change from Pre to Posttest		
Measures	*M*	*SD*	*P*	*M*	*SD*	*p*
T	2.57	6.09	0.048*	-0.05	4.03	0.86
O	-1.10	5.49	0.56	-1.35	4.97	0.22
A	-2.05	4.72	0.08	0.50	3.85	0.72
S	0.48	6.31	0.70	-0.90	5.88	0.37
C	-0.43	3.92	0.56	0.40	5.11	0.59
I	-2.90	7.62	0.26	1.05	4.88	0.58
M	-0.90	5.98	0.45	-0.95	4.82	0.45

Note. T = Truth-seeking, O = Open-mindedness, A = Analyticity, S = Systematicity; C = CT Self-Confidence; I = Inquisitiveness; M = Maturity.

* $p<0.05$

Table 5b Continued: *Results from Wilcoxon Sign Rank Test: Determination of Significant Gains within Each CCTDI Subscale for Each Participating Section*

	Experimental 5 (n = 19) Change from Pre to Posttest			Traditional 1 (n = 20) Change from Pre to Posttest		
Measures	*M*	*SD*	*p*	*M*	*SD*	*p*
T	-2.37	3.45	0.01*****	-0.45	4.20	0.62
O	-1.32	4.46	0.27	-0.05	4.88	0.74
A	-1.211	4.12	0.35	-1.10	3.11	0.14
S	-0.316	3.33	0.64	0.10	4.38	0.95
C	1.53	4.99	0.21	0.15	4.00	0.88
I	-1.37	4.07	0.20	-2.10	3.66	0.025*****
M	-1.63	5.79	0.31	-0.70	3.83	0.43

Note. T = Truth-seeking, O = Open-mindedness, A = Analyticity, S = Systematicity; C = CT Self-Confidence; I = Inquisitiveness; M = Maturity.

* $p<0.05$

	Traditional 2 (n = 83) Change from Pre to Posttest			Traditional 3 (n = 13) Change from Pre to Posttest		
Measures	*M*	*SD*	*p*	*M*	*SD*	*p*
T	-0.05	4.99	0.93	-0.08	4.65	0.73
O	-0.71	5.65	0.06******	-1.23	3.19	0.22
A	-0.65	5.10	0.35	0.85	3.16	0.43
S	-0.36	5.18	0.43	-0.38	5.19	0.55
C	-0.10	5.44	0.86	1.85	7.49	0.64
I	-0.71	5.24	0.15	0.46	6.83	0.72
M	-1.16	6.00	0.07******	-0.92	4.66	0.42

Note. T = Truth-seeking, O = Open-mindedness, A = Analyticity, S = Systematicity; C = CT Self-Confidence; I = Inquisitiveness; M = Maturity.

** $p<0.10$

Table 5b Coninuted: *Results from Wilcoxon Sign Rank Test: Determination of Significant Gains within Each CCTDI Subscale for Each Participating Section*

	Traditional 4 (*n* = 19) Change from Pre to Posttest			Traditional 5 (*n* = 62) Change from Pre to Posttest		
Measures	*M*	*SD*	*p*	*M*	*SD*	*p*
T	-0.47	3.61	0.71	-0.07	6.29	0.97
O	-1.05	3.72	0.27	-0.86	4.56	0.24
A	-0.68	5.37	0.78	1.05	4.74	0.08**
S	-0.053	4.14	0.88	0.55	5.36	0.37
C	-0.11	5.29	0.59	-0.02	4.95	0.79
I	-2.68	5.90	0.04*	-1.53	5.64	0.07**
M	1.79	3.44	0.03*	-0.68	6.84	0.45

Note. T = Truth-seeking, O = Open-mindedness, A = Analyticity, S = Systematicity; C = CT Self-Confidence; I = Inquisitiveness; M = Maturity.

* p<0.05, ** p<0.10

	Traditional 6 (*n* = 58) Change from Pre to Posttest		
Measures	*M*	*SD*	*p*
T	0.57	5.41	0.29
O	-1.31	5.64	0.17
A	-0.29	4.65	0.96
S	0.17	5.10	0.81
C	-0.05	6.01	0.89
I	-0.35	5.57	0.63
M	-0.83	6.69	0.64

Note. T = Truth-seeking, O = Open-mindedness, A = Analyticity, S = Systematicity; C = CT Self-Confidence; I = Inquisitiveness; M = Maturity.

FIGURES

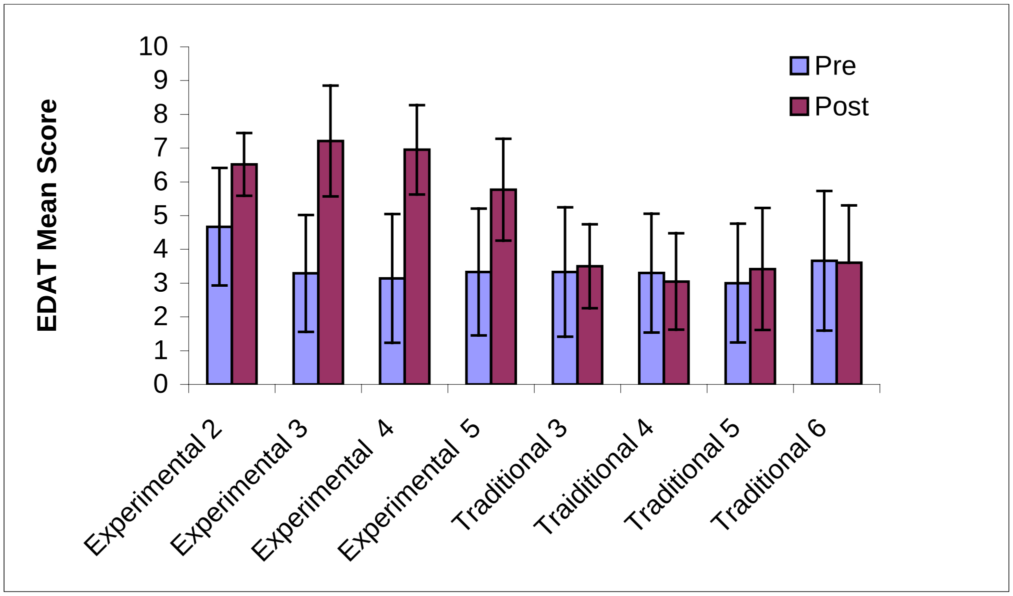
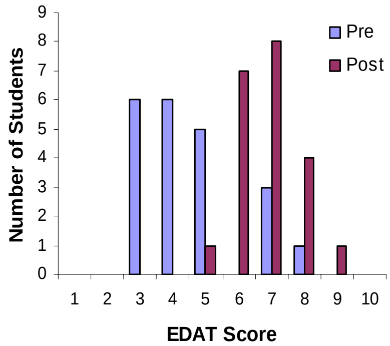
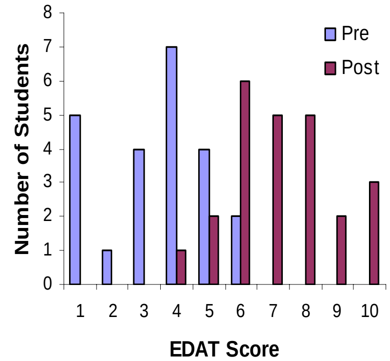


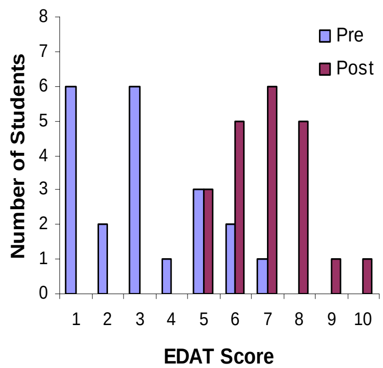
Figure 1. Distribution of Pre and Post EDAT Means +/- Standard Deviation for each Group. Gains in average EDAT scores for the experimental groups are statistically significant ($p<0.001$) as determined by the One-sample Wilcoxon sign rank test. Traditional groups did not make statistically significant gains.



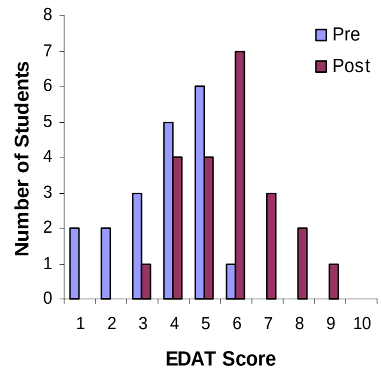
A)



B)



C)



D)

Figure 2. (A-D) Frequency distribution of student pre and post test EDAT scores for experimental teaching sections for Introductory Biology 104. The x-axis of the graphs show the EDAT score and the y-axis shows the number of students with that score. (A) Experimental group 2. (B) Experimental group 3. (C) Experimental group 4. (D) Experimental group 5.

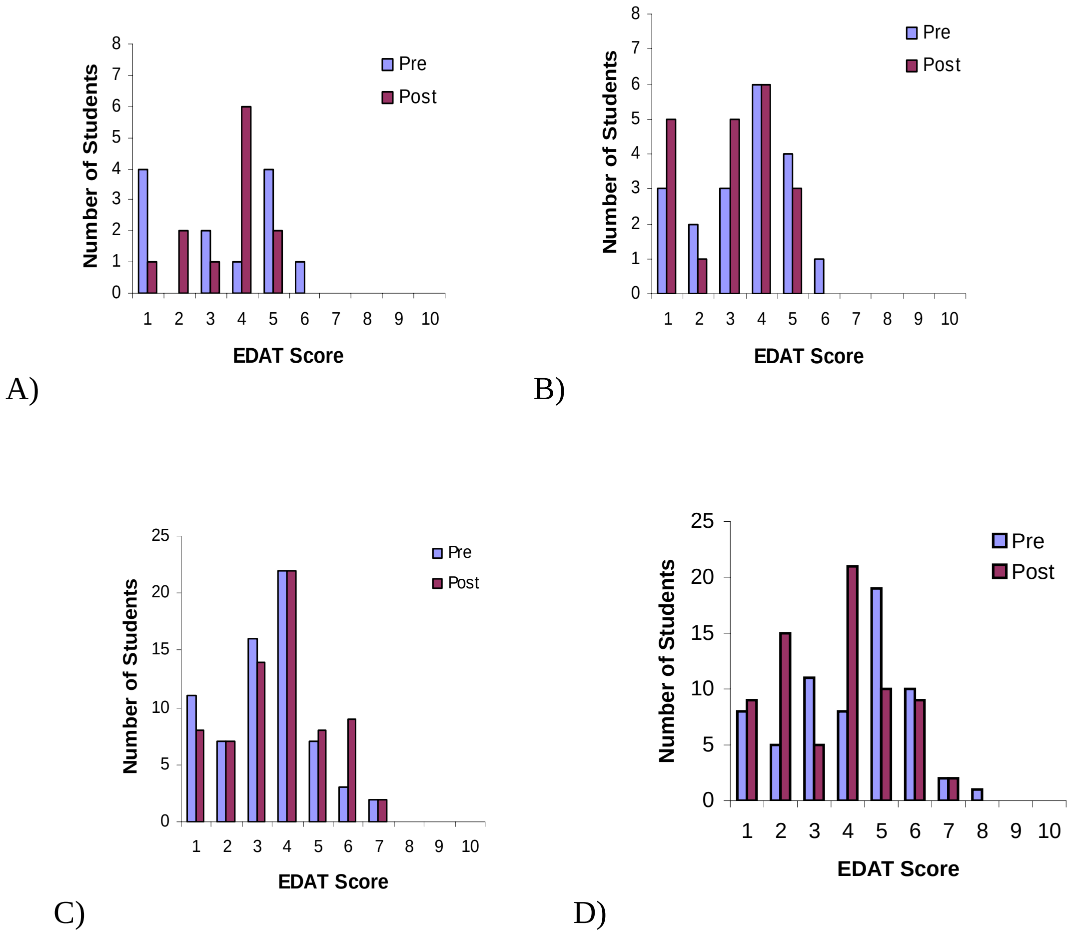
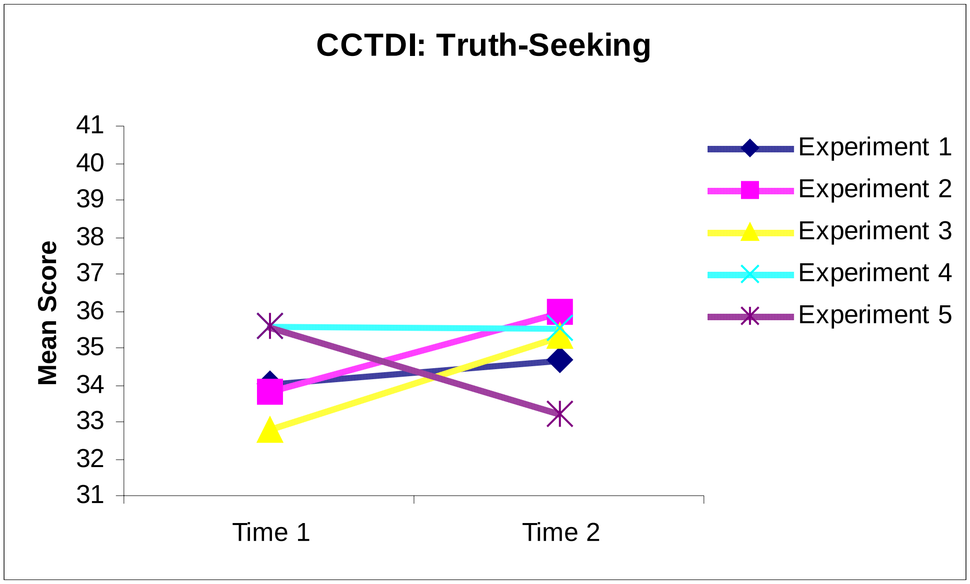
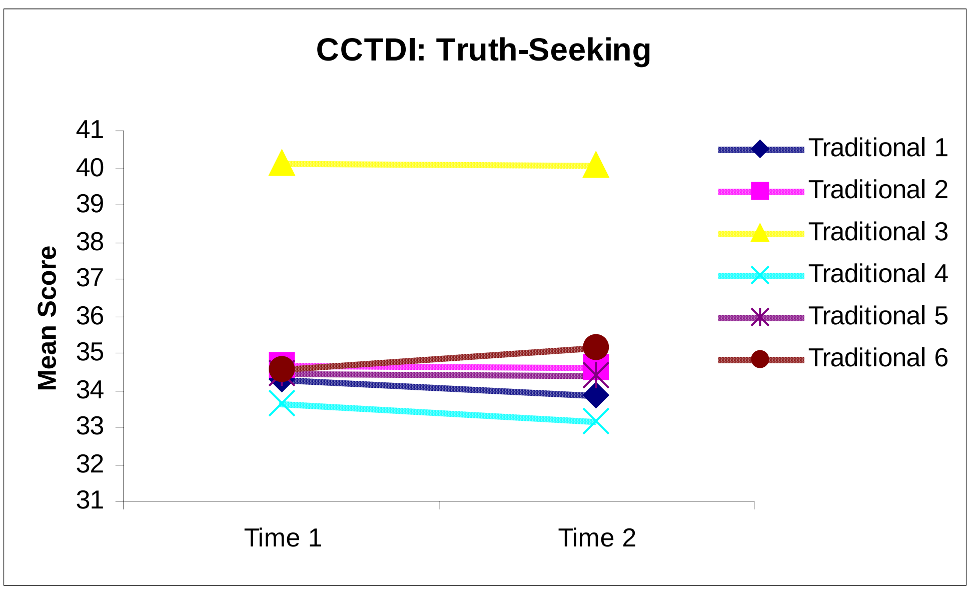


Figure 3. (A-D) Frequency distribution of student pre and post test EDAT scores for traditional teaching sections for Introductory Biology 104. The x-axis of the graphs show the EDAT score and the y-axis shows the number of students with that score. (A) Traditional group 3. (B) Traditional group 4. (C) Traditional group 5. (D) Traditional group 6.

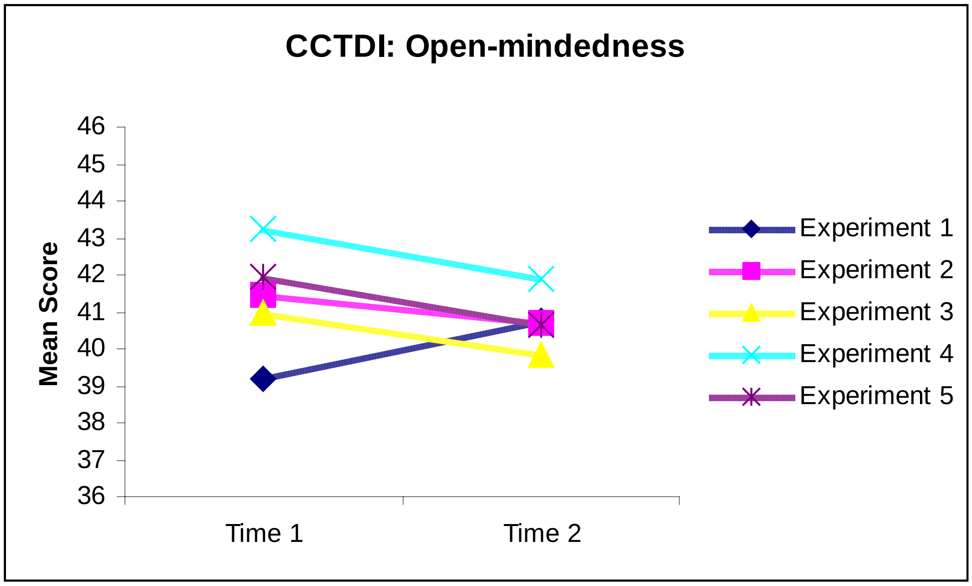


(A)

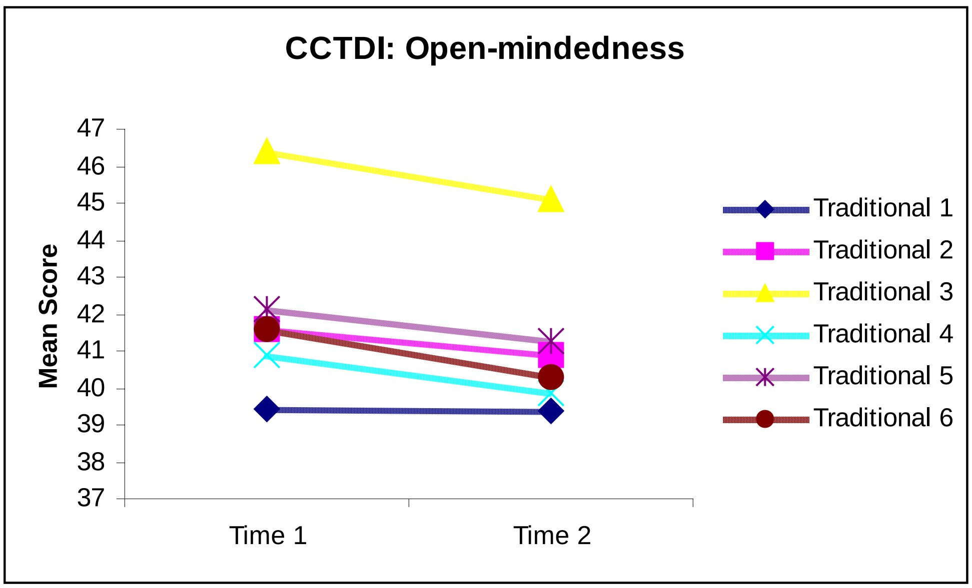


(B)

Figure 4. (A-B) Trends in the students' change in Critical Thinking Disposition for Truth-seeking. (A) Experimental groups. (B) Traditional groups.

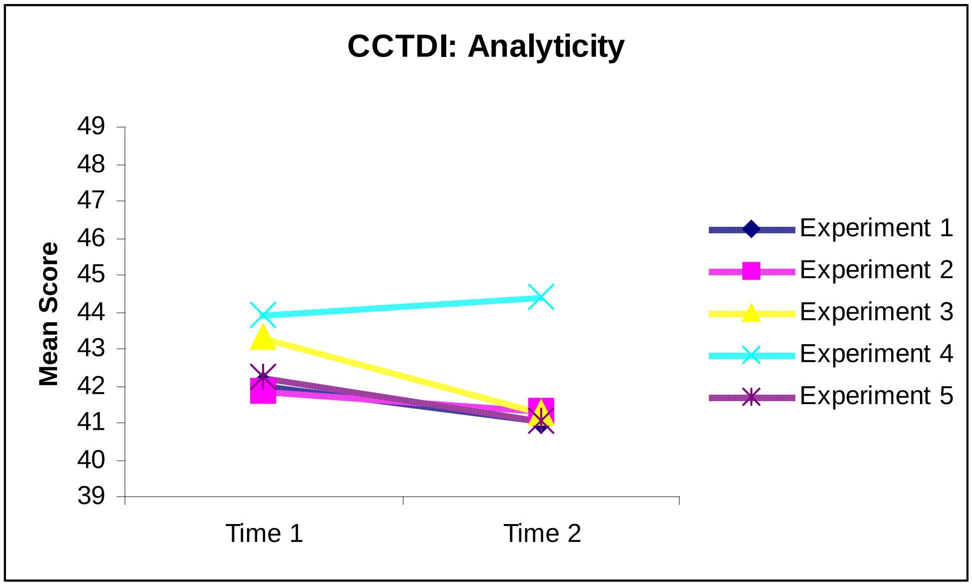


(A)

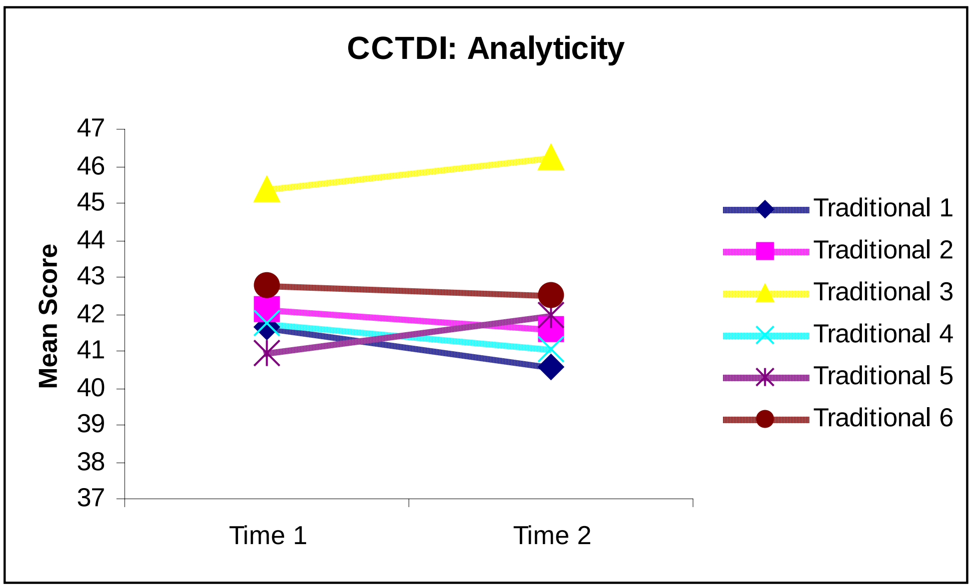


(B)

Figure 5. (A-B) Trends in the students' change in Critical Thinking Disposition for Open-mindedness. (A) Experimental groups. (B) Traditional groups.

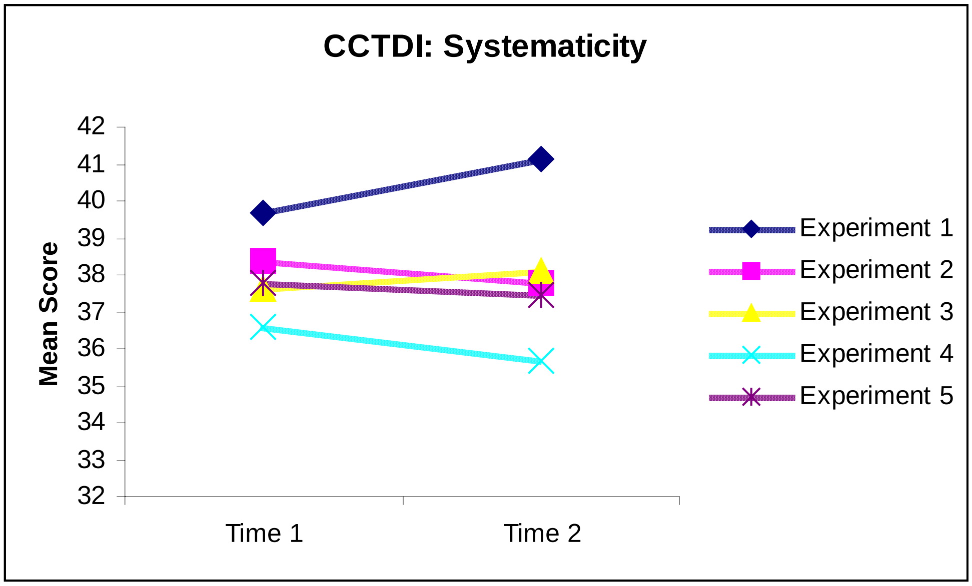


(A)

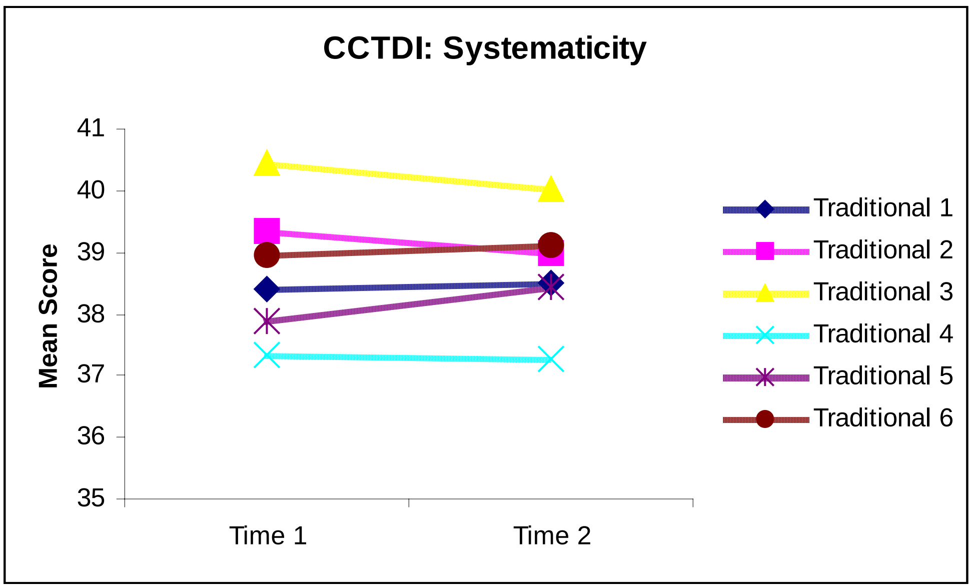


(B)

Figure 6. (A-B) Trends in the students' change in Critical Thinking Disposition for Analyticity. (A) Experimental groups. (B) Traditional groups.

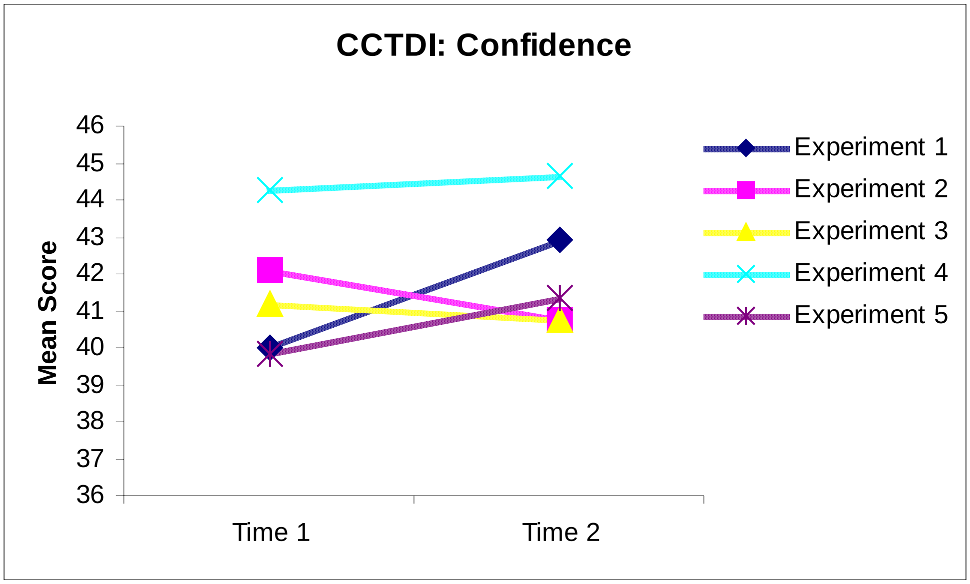


(A)

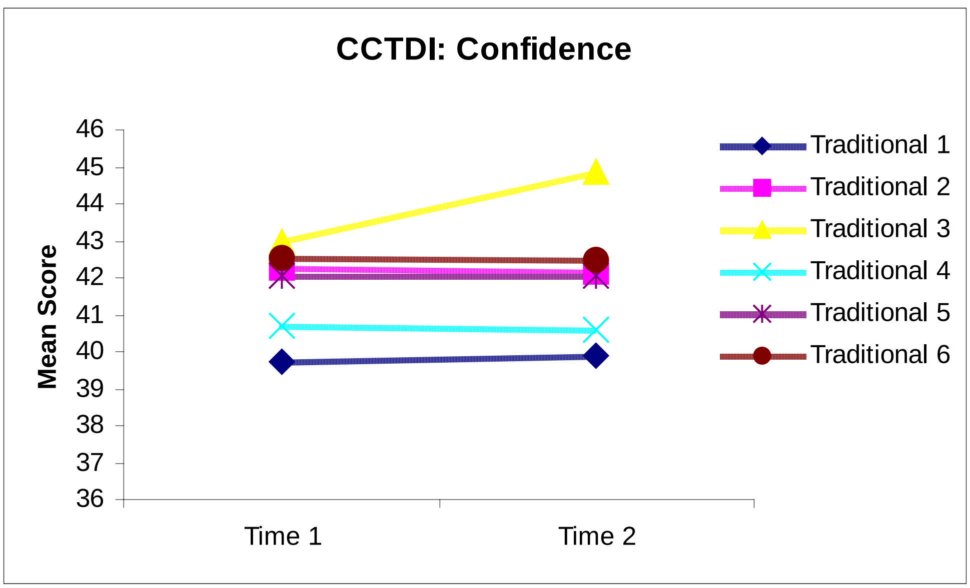


(B)

Figure 7. (A-B) Trends in the students' change in Critical Thinking Disposition for Systematicity. (A) Experimental groups. (B) Traditional groups.

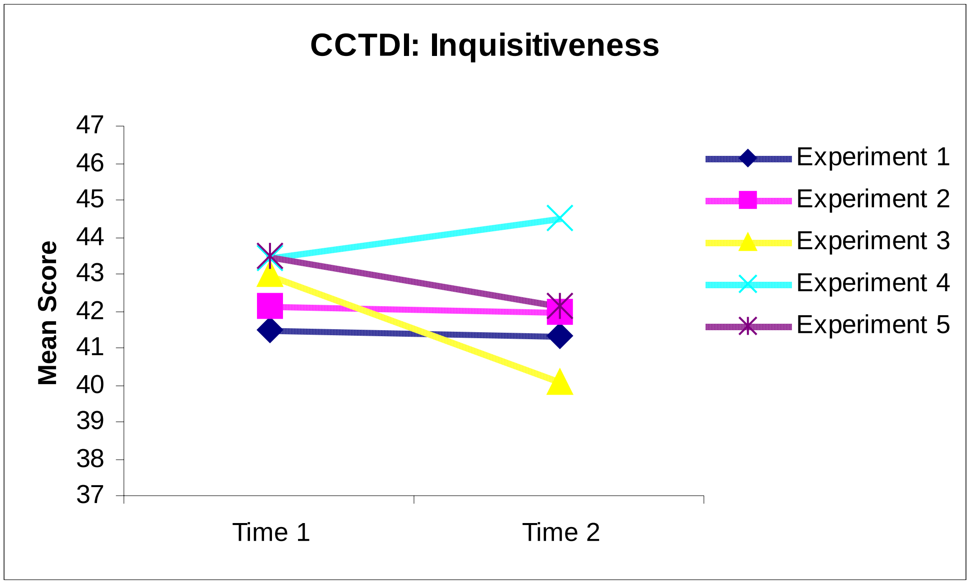


(A)

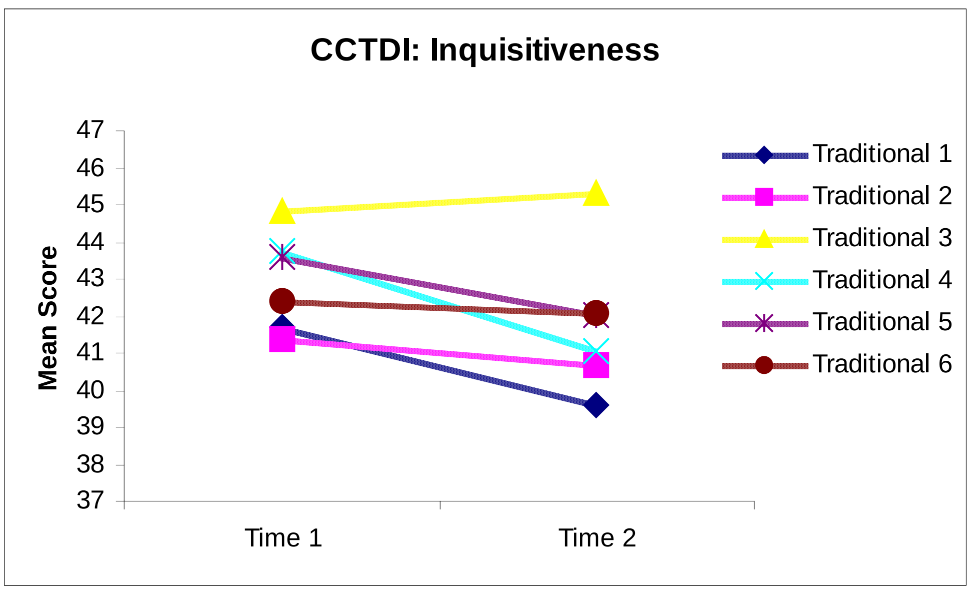


(B)

Figure 8. (A-B) Trends in the students' change in Critical Thinking Disposition for Confidence. (A) Experimental groups. (B) Traditional groups.

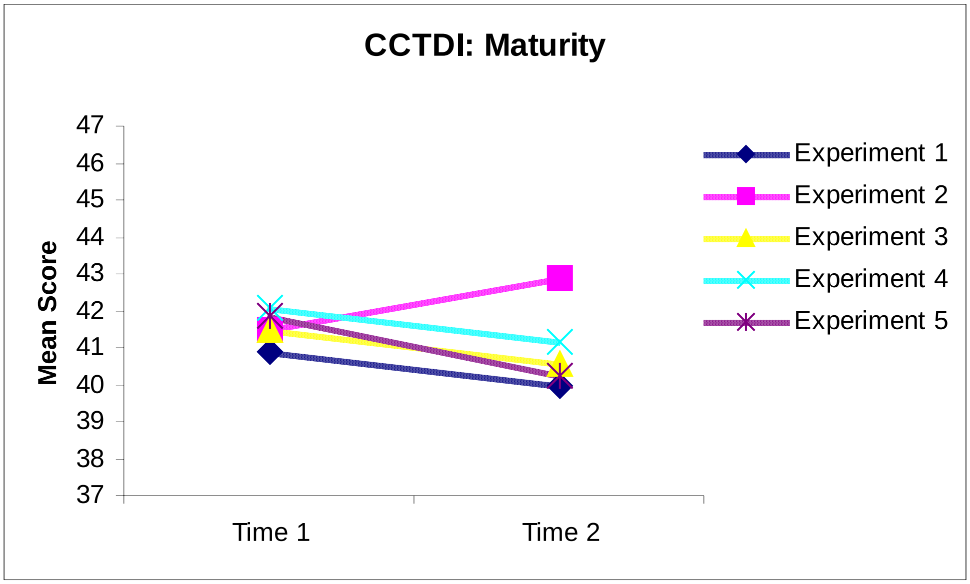


(A)

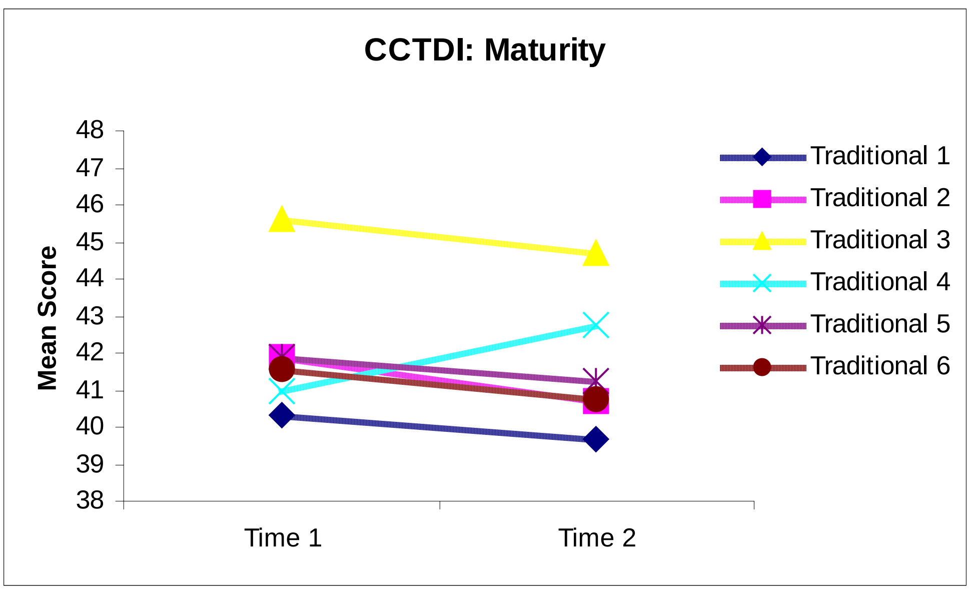


(B)

Figure 9. (A-B) Trends in the students' change in Critical Thinking Disposition for Inquisitiveness. (A) Experimental groups. (B) Traditional groups.



(A)



(B)

Figure 10. (A-B) Trends in the students' change in Critical Thinking Disposition for Maturity. (A) Experimental groups. (B) Traditional groups.

APPENDIX A

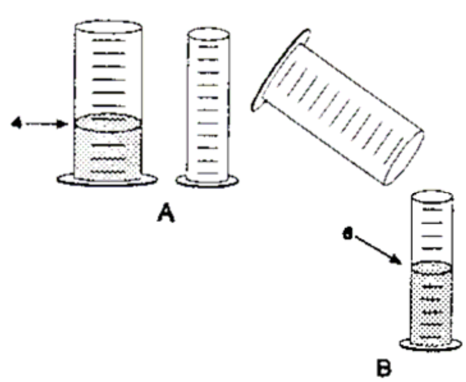
Classroom Scientific Reasoning Test Sample Question

5. To the right are drawings of a wide and a narrow cylinder. The cylinders have equally spaced marks on them. Water is poured into the wide cylinder up to the 4th mark (see A). This water rises to the 6th mark when poured into the narrow cylinder (see B).
Both cylinders are emptied (not shown) and water is poured into the wide cylinder up to the 6th mark. *How high would this water rise if it were poured into the empty narrow cylinder?*

a. to about 8
b. to about 9
c. to about 10
d. to about 12
e. none of these answers is correct

6. *because*

a. the answer can not be determined with the information given.
b. it went up 2 more before, so it will go up 2 more again.
c. it goes up 3 in the narrow for every 2 in the wide.
d. the second cylinder is narrower.
e. one must actually pour the water and observe to find out.



Revised Edition: August 2000 by Anton E. Lawson, Arizona State University. Based on: Lawson, A.E. 1978. Development and validation of the classroom test of formal reasoning. *Journal of Research in Science Teaching*, 15(1): 11-24.

APPENDIX B

California Critical Thinking Skills Test (CCTST)

Reasoning and Critical Thinking Skills Sample Test Items

Instructions: Select the best choice from among those offered.

Given the importance of critical thinking to our democracy, our economy, and our lives in a pluralistic, global community, we hope that you get most, if not all, of these right.

Sample Reasoning Skills Item #1: Using the phone at her desk, Sylvia in Corporate Sales consistently generates a very steady $1500 per hour in gross revenue for her firm. After all of her firm's costs have been subtracted, Sylvia's sales amount to $100 in bottom line (net) profits every 15 minutes. At 10:00 a.m. one day the desk phone Sylvia uses to maker her sales calls breaks. Without the phone Sylvia cannot make any sales. Assume that Sylvia's regular schedule is to begin making sales calls at 8:00 a.m. Assume she works the phone for four hours, takes a one hour lunch exactly at noon, and then returns promptly to her desk for four more hours of afternoon sales. Sylvia loves her work and the broken phone is keeping her from it. If necessary she will try to repair the phone herself. Which of the following options would be in the best interest of Sylvia's firm to remedy the broken phone problem?

A = Use Ed's Phone Repair Shop down the street. Ed can replace Sylvia's phone by 10:30 a.m. Ed will charge the firm $500.

B = Assign Sylvia to a different project until her phone can be replaced with one from the firm's current inventory. Replacing the phone is handled by the night shift.

C = Authorize Sylvia to buy a new phone during her lunch hour for $75 knowing she can plug it in and have it working within a few minutes after she gets back to her desk at 1:00 p.m.

D = Ask Sylvia to try to repair her phone herself. She will probably complete the repair by 2:00 p.m.; or maybe later.

Facione (1990, 1998) in http://www.insightassessment.com/test-cctst2k.html

APPENDIX C

EDAT ASSESSMENT INSTRUMENT

Name_____________________________ Date_______________________

P#_________________________ Course and Instructor _____________________

Pretest Prompt: Advertisements for an herbal product, ginseng, claim that it promotes endurance. To determine if the claim is fraudulent and prior to accepting this claim, what type of evidence would you like to see? Provide details of an investigative design.

Posttest Prompt: The claim has been made that women may be able to achieve significant improvements in memory by taking iron supplements. To determine if the claim is fraudulent and prior to accepting this claim, what type of evidence would you like to see? Provide details of an investigative design.

Experimental Design Ability Test Scoring Rubric:

Student response to the problem posed in the prompt, the student demonstrates:

__1. Recognition that an experiment can be done to test the claim (*vs.* simply reading the product label).

__2. Identification of what variable is manipulated (independent variable is ginseng *vs.* something else).

__3. Identification of what variable is measured (*e.g.,* how far subjects run will be measure of endurance).

__4. Description of how dependent variable is measured (*e.g.,* how far subjects run will be measure of endurance).

__5. Realization that there is <u>one other</u> variable that must be held constant (*vs.* no mention).

__6. Realization that there are <u>many</u> variables that must be held constant (*vs.* only one or no mention).

__7. Understanding that the larger the sample size or number of subjects, the better the data.

__8. Understanding that the experiment needs to be repeated.

__9. Understanding of the placebo effect (subjects do not know if they were given ginseng or a sugar pill).

__10. Awareness that one can never prove a hypothesis, that one can never be 100% sure, that there might be another experiment that could be done that would disprove the hypothesis, that there are possible sources of error, that there are limits to generalizing the conclusions (credit for any of these).

APPENDIX D

Views on Science-Technology-Society (VOSTS) Sample Question

Actual VOSTS question
Defining science is difficult because science is complex and does many things. But MAINLY science is: Your position, basically: (Please read from A to K, and then choose one) A. a study of fields such as biology, chemistry, and physics B. a body of knowledge, such as principles, laws and theories, which explain the world around us (matter, energy and life) C. exploring the unknown and discovering new things about our world and universe and how they work D. carrying out experiments to solve problems of interest about the world around us E. inventing or designing things (for example, artificial hearts, computers, space vehicles) F. finding and using knowledge to make this world a better place to live in (for example, curing diseases, solving pollution and improving agriculture) G. an organization of people (called scientists) who have ideas and techniques for discovering new knowledge H. No one can define science I. I don't understand J. I don't know enough about this subject to make a choice K. None of these choices fits my basic veiwpoint

Aikenhead & Ryan (1992) in
http://www.flaguide.org/tools/attitude/views_about_sciences.php

APPENDIX E

Views About Science Survey (VASS) Sample Question

Actual VASS Question
Learning physics requires: (a). serious effort. (b). a special talent. What would each one of the five choices mean? 1. Mostly (a), rarely (b): Learning physics requires mostly a serious effort and rarely a special talent (or mainly the former and hardly ever the latter). 2. More (a) than (b): Learning physics requires more a serious effort than a special talent. 3. Equally (a) & (b): Learning physics requires as much a serious effort as a special talent. 4. More (b) than (a): Learning physics requires more a special talent than a serious effort. 5. Mostly (b), rarely (a): Learning physics requires mostly a special talent and rarely a serious effort (or mainly the former and hardly ever the latter).

Halloun and Hestenes (1998) in
http://www.flaguide.org/tools/attitude/views_about_sciences.php

APPENDIX F

California Critical Thinking Disposition Inventory (CCTDI)

Sample Reasoning Motivation and Disposition Items

Consider the following 10 statements about beliefs, opinions, values, and preferences.
Decide whether you agree or disagree with each one. Remember that since you are being asked about your own beliefs, opinions, values, and preferences, there really is no "right" or "wrong" response. The answer is whatever you say it is for you.

You can indicate the extent of your affirmation or rejection of each statement by giving each one a point value where as follows.

6 = Agree Strongly
5 = Agree
4 = Agree Marginally
3 = Disagree Marginally
2 = Disagree
1 = Disagree Strongly

1. I hate talk-radio hosts because they shout out their views without really listening to the other side.

2. I won't let what scientists might say weaken my core beliefs.

3. I prefer jobs where the supervisor says exactly what to do, and exactly when and how to do it.

4. It's important to me to figure out what people really mean by what they say.

5. Don't kid yourself, changing your mind is a sign of weakness.

6. I always do better in jobs where I'm expected to think things out for myself.

7. If I wanted to persuade someone of something, I wouldn't stop talking until the person gave up.

8. My friends expect me to be able to figure out a smart way to deal with all kinds of problems.

9. For me the best way to make decisions is to go with my gut feelings.

10. I hold off making decisions until I've thought through my options.

Facione and Facione (1992) in http://www.insightassessment.com/Sample%20Test2.html