Running Head: CRITICAL THINKING RESEARCH

Implementing Socratic Questioning and Inquiry Training to Help Students Achieve Higher Level Critical Thinking Skills

Introduction

The world is rapidly changing around us. To complete globally, today's students must have critical thinking skills. They must be able to understand the full context of a situation, be able to analyze it, and find solutions to problems. These skills need to be acquired before the students hit the job market. Students need to be achieving at higher levels than ever before in school. "Critical thinking is universally recognized as important in schooling at all levels," (Foundation for Critical Thinking, 2009). However, after the implementation of No Child Left Behind (NCLB) standardized tests have become the means of measuring student achievement. The Illinois Standards Achievement Test (ISAT), and the Measure of Academic Progress (MAP) are two of the tests used by teachers and administration to check student progress. ISAT is mostly multiple choice questions with two short responses and one extended response question for math. Writing is only evaluated in third, fifth, sixth, and eighth grades. MAP testing is computer-based and therefore is multiple choice questions. Although the MAP tests do adjust to the students' level, as they answer the questions correctly subsequent questions become harder and as they answer the questions incorrectly subsequent questions become easier, it has no written component. Both ISAT and MAP testing offer very little chance for students to think critically.

I teach at Kenyon Woods Middle School in South Elgin, Illinois, which is part of School District U-46 based in Elgin, Illinois. Within the context of my classroom, I will implement different methods of instruction to help student achieve higher level thinking skills. Although the students are involved in an inquiry-based science program, most of the investigations are not asking as many probing questions requiring the students to synthesize and evaluate the situations. The students are not applying what they have learned from the investigation into real-world applications or into their everyday lives. Although the students participate in a variety of activities, only a small percentage leads the students to applying the concepts learned to other scenarios or creating their own projects. If the students can make the connections into their everyday lives, I believe there will be more engagement in the activities. If students analyze the data and apply the results, they will start to think more critically, which will lead them to make better predictions, create their own projects, or form their own opinions. Therefore, my research question is “Is there a difference in the mean test scores between inquiry training and Socratic questioning?” These methods are important to my research because I want to improve students’ critical thinking skills. By implementing these strategies, I hope to have the students thinking at a deeper level to understand the world around them.

Background

As far back as Socrates, critical thinking skills have always been a part of society. Socrates “established the importance of asking deep questions that probe profoundly into thinking before we accept ideas as worthy of belief” (Foundation, 2009). Plato and Aristotle used the practice of deeper questioning to understand the world around them. Throughout the Middle Ages, the Renaissance, and up to present-day, critical thinking has been a topic of interest to many scientists and politicians. The Declaration of Independence was produced out of developing critical reasoning. Copernicus, Galileo, Kepler, Sir Isaac Newton, Darwin, and Freud, brought critical thinking into their fields of study from the stars to the unconscious mind. Critical thinking skills have been brought into every field of study. Being able to look deeper into a problem and understand what is happening at a deeper level is the key to being able to think deeper and analyze a situation.

John Dewey was one of the founders of progressive education. “The primary thesis inherent in much of Dewey's work is that children, through their play, emulate and experiment with activities in the social milieu, thereby developing practical skills, academic skills, and critical thinking skills which they then continue to apply to the society in which they live and work” (Braundy, 2004). Dewey believed students should be involved with curriculum that was both integrated with other subjects and involve active experiences. He believed the teacher was the guide to the students learning not the task-master, making the students responsible for their learning. Discrepant events and science mysteries actively engage the students in thinking more critically.

Benjamin Bloom believed that all students could learn if given quality instruction. “Critical elements of this quality of instruction were (1) clearly communicating the learning expectations; (2) giving students specific feedback as to their progress in achieving them; and (3) providing additional time and help as needed by students” (Anderson, 2003). Bloom identified the cognitive, affective, and psychomotor domains for learning. The cognitive domain involves acquiring knowledge and intellectual skills. The affective domain involves feelings and emotions. The psychomotor domain involves physical or manual skills. Bloom created six different levels of learning for the cognitive domain. These are more commonly known as Bloom’s Taxonomy. These include knowledge, comprehension, application, analysis, synthesis, and evaluation. As a students moves through the levels they become better critical thinkers. The three highest levels, analysis, synthesis, and evaluation require the students to think about situations and troubleshoot, design, or explain and justify their answers. The use of Socratic questions can also fall into these same taxonomy levels.

Stillings et al. (1999) assessed the effects of inquiry based science classrooms compared to a more traditional lecture style classroom. They believed students in an inquiry based classroom would be able to increase their scientific reasoning abilities. Students at Hampshire College involved in the study were assessed by using a pre-test and a post-test. The pre-test was administered at the orientation session and the post-test was given during the last two weeks of the term. Another college was involved as a comparison college. The data was broken down into four groups: Students in inquiry science classes, students not enrolled in science classes, students enrolled in traditional biology classes, and students enrolled in reform biology classes. Students in inquiry-based science classes improved their critical thinking scores more than the other three groups. Although the study was completed with first year college students and I teach seventh grade students, I believe this is valuable research. The study involved students in inquiry-based science classrooms. I also teach students using an inquiry-based curriculum.

Methodology

For this study, the target population is all seventh grade science students at Kenyon Woods Middle School. Participants in the study are limited to 21 students in my A block class and the 26 students in my E block class. Twenty-two boys and 22 girls are participating. This is a Cluster Random Sample. The students are divided into five different classes and each class is representative of the population. Two clusters (classes) chosen are also representative of the population ethnicities, with just slightly lower grades than the population, sample mean = 83.5%, population mean = 87%. Individual students of the two clusters chosen will be held accountable for the information and their scores will be used. Of these 47 students, three will not be participating in the study. The accessible population is 125 students in my five science classes, 64 students are female and 61 students are male.

The strategies implemented are Inquiry Training and Socratic Questioning, the independent variables. The first strategy, Inquiry Training, will involve the use of discrepant events and science mysteries. Discrepant events are operationally defined as events that appear not to follow the rules of nature and should motivate students to ask questions and want to investigate further. Science mysteries are operationally defined as short scenarios that require students to think analytically to figure out the science behind the mystery. The second strategy, Socratic questioning, is operationally defined to help teachers assess student knowledge and help students analyze a concept by asking different levels of questions. I manipulate these variables among the two classes and administer a critical thinking assessment before the treatments begin. I implement a different strategy to each class; afterwards I assess the students with the same critical thinking assessment. I switch the treatments and again assess the students. This way I am replicating the research to help reach my conclusions. The discrepant events and science mysteries are implemented over a two week period. At least seven to eight total events or mysteries are implemented over the two week period. This requires some time out of the normal class schedule to implement. Socratic questions are imbedded within the lessons being taught and require no additional time. The science classes are involved in an inquiry-based science program, so both classes are conducting the same investigations and working in cooperative groups throughout the school year.

Critical thinking skills, dependent variable, are operationally defined as students being able to take the knowledge they have learned and analyze and evaluate it. The variable is measured by using the same Critical Thinking assessment three times, before any treatment and after each treatment. In the assessment students are given two scenarios they must form questions or reasons for, write a hypothesis, explain their reasoning, and make a prediction. This is a paper and pencil assessment with a rubric for each question. These are assessed using up to a three point scale for each part of the questions. Each question has three parts to it, A, B, and C. A total score of 14 is the highest score that can be achieved.

The null hypothesis states there is no difference in the mean test scores, and the alternative hypothesis states there is difference in the mean test scores.

$H_O: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 - \mu_2 \neq 0$

An alpha level of .05 for a two tail test will be used with a 95% confidence interval and degrees of freedom is 43.

Limitations to the study are the number of questions asked in each class, the amount of mysteries implemented, and because we are in a unit that involves live organisms, having the time to implement the strategies correctly and in a timely manner, while still keeping our organisms alive. Another limitation may the be assessment itself, because the students will take the same assessment three times. Are they getting better because of the strategies I implemented or because this is the third time they have taken the same assessment?

Results

Socratic questioning and inquiry training were implemented to both of my A and E block students. The students first took a pretest to assess the level of their critical thinking skills. Bock A received the Socratic questioning technique. Over the two week period every time questions arose I would ask the students another question to help them figure out the answers. The students were in class six times in that two week period, twice for 43 minutes and four times for 90 minutes each. It was during the 90 minute blocks that the Socratic Questioning seemed to really be implemented. E block received the inquiry training over the same two week period. At the beginning of each class, students were given a science mystery to solve. I saw these students the same amount of time as A block. During the 90 minute science blocks the students were exposed to two mysteries. This took about 15 to 20 minutes but the students really enjoyed them and asked to do more. Both A and E Block students then took the critical thinking assessment. The methods of treatment were then switched over the next two weeks. I implemented the same science mysteries for the students but because of the content being covered, the Socratic questions were different. The students were given their final critical thinking assessment at the end of the two week period. A dependent sample t-test was run. The paired samples compared the means of the critical thinking assessment scores following the implementation of Socratic Questioning and Inquiry Training. The statistical procedures were run under some assumptions. The following are assumptions for this study:

1. The population of differences for each distribution is normally distributed.
2. The sample differences are a random sample from the population of differences.

Descriptive and inferential statistics tables are included in Appendix A. Statistical Package for the Social Sciences (SPSS) was used to calculate the statistics. Forty-four students were involved in the study, which satisfies the central limit theorem for sample size of $n \geq 30$. The mean is used as the measure of central tendency for this study. The mean for Socratic Questioning (SQ) is 7.95 (SD 2.676). The skewness of SQ is calculated at -.010. This data shows that the skewness is in the normal range between -2.00 to +2.00, so the data does not show a problem with skewness and the assumption of the sample differences is normally distributed is met. Kurtosis for SQ is calculated at -.660. Since the values fall within the normal range of -2.00 to +2.00 there is not a problem with kurtosis which is Mesokurtic, meaning the data falls with in a normal bell curve.

The mean for Inquiry Training (IT) is 8.09 (SD 2.409). The skewness for IT is calculated at .076. This data shows that the skewness is in the normal range between -2.00 to +2.00, so the data does not show a problem with skewness and the assumption of the sample differences is normally distributed is met. Kurtosis for IT is calculated at -.646. Since the values fall within the normal range of -2.00 to +2.00 there is not a problem with kurtosis which is Mesokurtic, meaning the data falls with in a normal bell curve.

The paired samples correlations box shows a correlation value of .574 between the two groups. This is a moderate, positive correlation between the two groups. The correlation significance of $p < .001$ indicates there is chance that the two groups are related and that the data should be paired. The standard error of the mean is calculated at .356 meaning there is slightly more than a one-third of a point difference between the two groups.

The null hypothesis states there is no difference in the mean test scores, and the alternative hypothesis states there is difference in the mean test scores. (H_O: $\mu_1 - \mu_2 = 0$: H_1: $\mu_1 - \mu_2 \neq 0$). The calculated 95% confidence interval of the difference is -.853, .581. We are not confident that out of 100 trials 95 could contain the population parameter for the mean difference housed within the interval -.852, .581. The achieved critical value at the .05 level according to Gravetter & Wallnau's (2005) tabled critical value = 2.021 for a t-test with 43 degrees of freedom. The observed t-value = -.384 which does not exceed the critical value of 2.021. Since the confidence interval contains zero and the t-value does not exceed the critical value, we fail to reject the null hypothesis. (The null hypothesis equals zero and since there is a zero point between our 95% confidence interval of -.852, .581 the null is present meaning we fail to reject the null). Also, the alpha level was set at .05 and since the observed alpha level is $p = .703 > .05$ there is no statistical significance for this study therefore there is no effect size to report. The power for the dependent t-test = .0575 meaning there is a 6% probability of achieving statically significant results. Therefore those are not high probabilities rates and it backs the data that we fail to reject the null and we have not protected adequately against a type II error.

The difference between the two groups has a mean of -.14. The skewness is within the normal range of -2.00 to +2.00 at .082 meaning there is difference is normally distributed. Kurtosis for the difference between the two groups is also within the normal range of -2.00 to +2.00 at -.195; therefore, kurtosis which is Mesokurtic, meaning the data falls with in a normal bell curve. Looking at the box plot in Appendix A we can see the mean difference between SQ and IT is virtually zero with a range from -5 to 5.

Conclusions

Overall, there is little difference between the means of Socratic Questioning and Inquiry Training methods of instruction. There was not statistical significance in this study and I failed to reject the null hypothesis. This does not mean that the study had no value. It proved to me that either method of teaching helps students improve their critical thinking skills almost equally. The overall mean did improve from the pretest which had a mean of 5.93 to SQ which had a mean of 7.95 and IT which had a mean of 8.09 (See appendix B). The increase of almost 2 full points on a 14 point scale is about at 14% increase in the scores. While this may not seem like a big increase, imagine if all test scores including MAP and ISAT testing increased by 14%. A student in the 60%tile could increase the score to the 74%tile for MAP testing. It may also be the difference in going from below to meeting or from meeting to exceeding on ISAT testing. This could have implications for many schools that are on academic warning.

I will need to further study both methods of teaching over a longer period of time. I will then compare both classes against their pretest and two post tests to see if one method improved the scores more over the other method. I believe by implementing both strategies for longer periods of time, students will increase their critical thinking skills.

Appendix A

Frequencies

Statistics

		Socratic Questioning	Inquiry Training
N	Valid	44	44
	Missing	0	0
Mean		7.95	8.09
Std. Deviation		2.676	2.409
Skewness		-.010	.076
Std. Error of Skewness		.357	.357
Kurtosis		-.660	-.646
Std. Error of Kurtosis		.702	.702

Frequency Tables

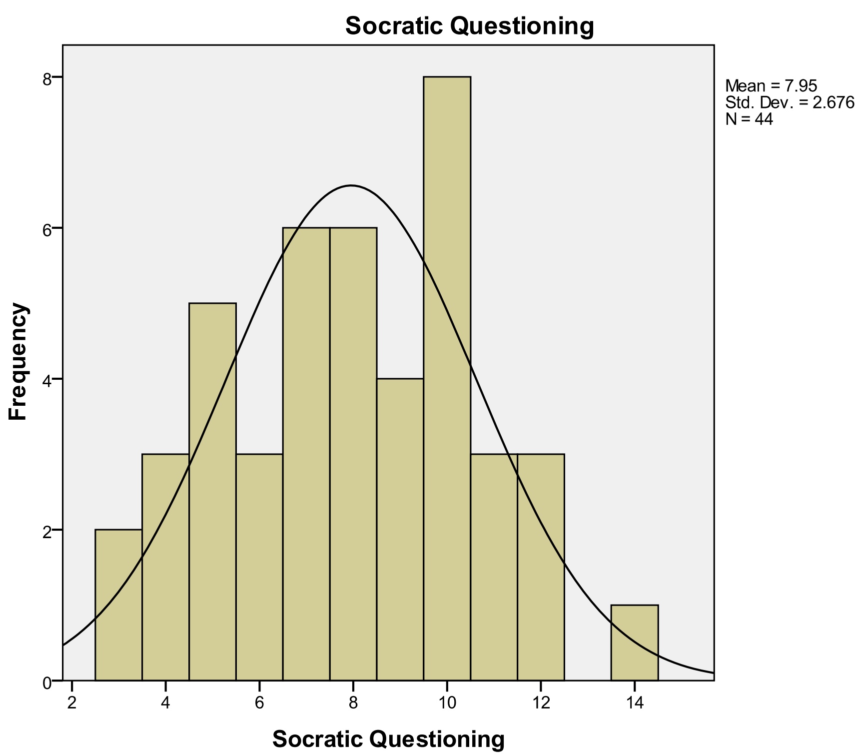
Socratic Questioning

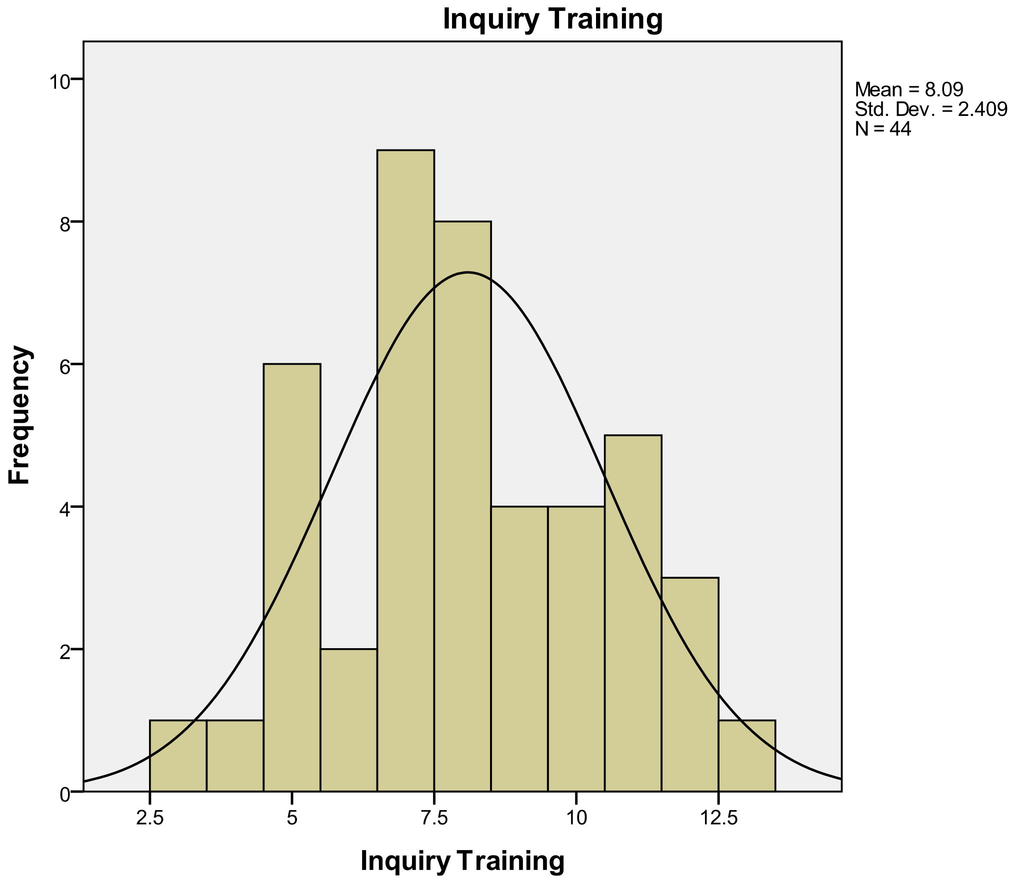
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3	2	4.5	4.5	4.5
	4	3	6.8	6.8	11.4
	5	5	11.4	11.4	22.7
	6	3	6.8	6.8	29.5
	7	6	13.6	13.6	43.2
	8	6	13.6	13.6	56.8
	9	4	9.1	9.1	65.9
	10	8	18.2	18.2	84.1
	11	3	6.8	6.8	90.9
	12	3	6.8	6.8	97.7
	14	1	2.3	2.3	100.0
	Total	44	100.0	100.0	

Inquiry Training

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3	1	2.3	2.3	2.3
	4	1	2.3	2.3	4.5
	5	6	13.6	13.6	18.2
	6	2	4.5	4.5	22.7
	7	9	20.5	20.5	43.2
	8	8	18.2	18.2	61.4
	9	4	9.1	9.1	70.5
	10	4	9.1	9.1	79.5
	11	5	11.4	11.4	90.9
	12	3	6.8	6.8	97.7
	13	1	2.3	2.3	100.0
	Total	44	100.0	100.0	

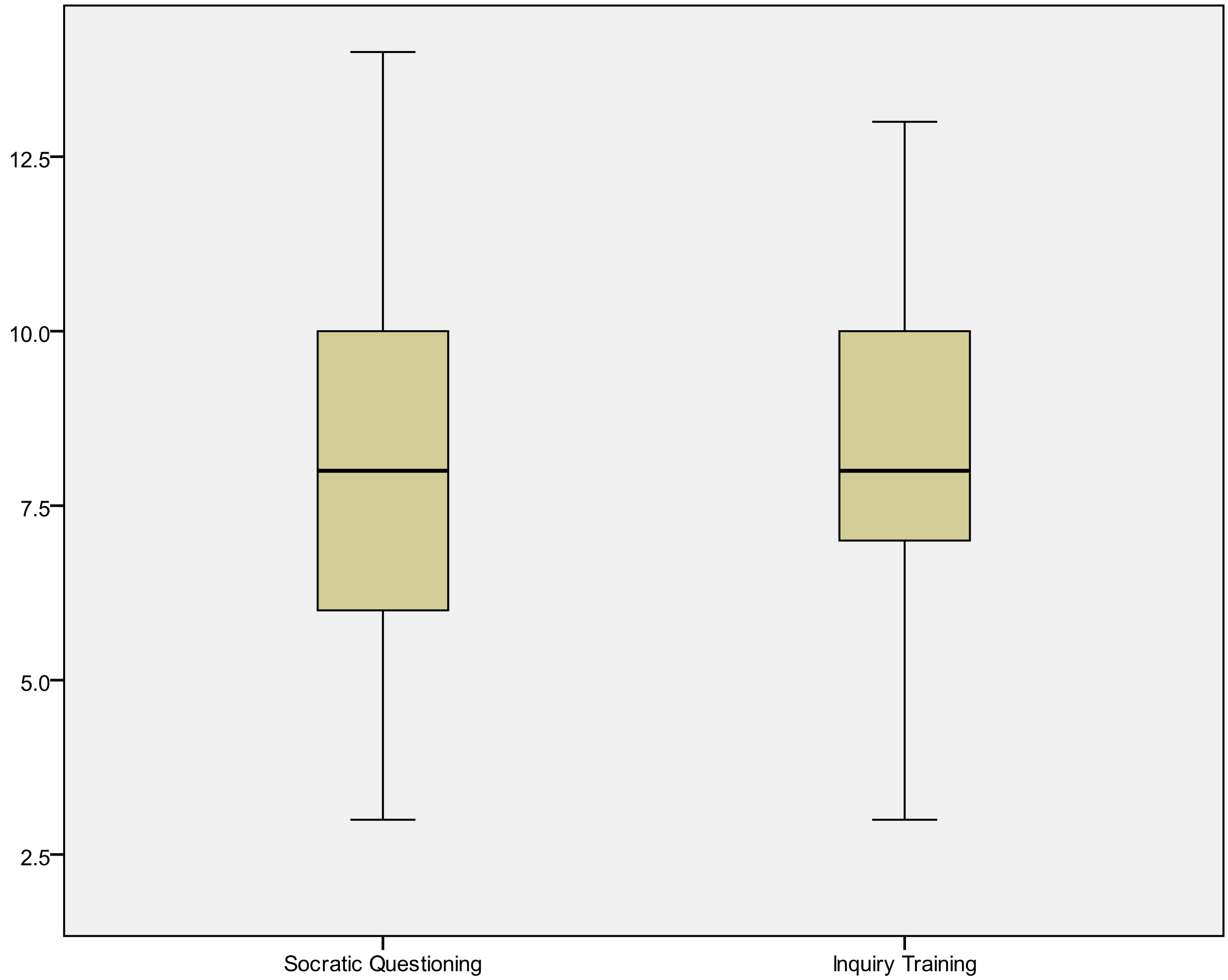
Histograms





Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Socratic Questioning	44	100.0%	0	.0%	44	100.0%
Inquiry Training	44	100.0%	0	.0%	44	100.0%



T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Socratic Questioning	7.95	44	2.676	.403
	Inquiry Training	8.09	44	2.409	.363

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Socratic Questioning & Inquiry Training	44	.574	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Socratic Questioning - Inquiry Training	-.136	2.358	.356	-.853	.581	-.384	43	.703

Frequencies

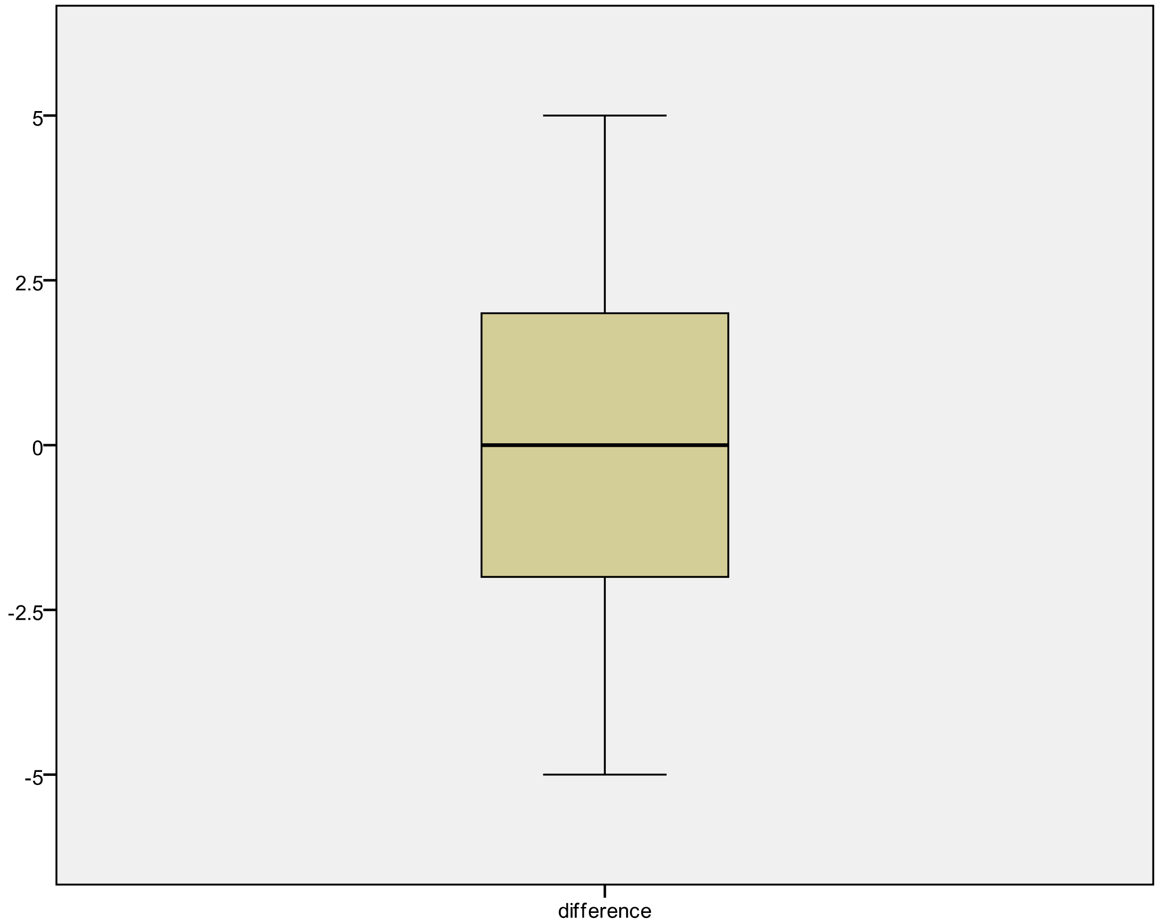
difference

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	-5	1	2.3	2.3	2.3
	-4	4	9.1	9.1	11.4
	-3	1	2.3	2.3	13.6
	-2	6	13.6	13.6	27.3
	-1	6	13.6	13.6	40.9
	0	12	27.3	27.3	68.2
	1	2	4.5	4.5	72.7
	2	6	13.6	13.6	86.4
	3	4	9.1	9.1	95.5
	5	2	4.5	4.5	100.0
	Total	44	100.0	100.0	

Explore

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
difference	44	100.0%	0	.0%	44	100.0%



Report

Descriptive Statistics

M1	SD1	n1	M2	SD2	n2
7.950	2.676	44	8.090	2.409	44

Effect Sizes for Standardized Differences Between Means and Power

Glass Delta	Hedges g	Cohens d	Power
-.0581	-.0545	-.0550	.0575

Appendix B

Descriptives

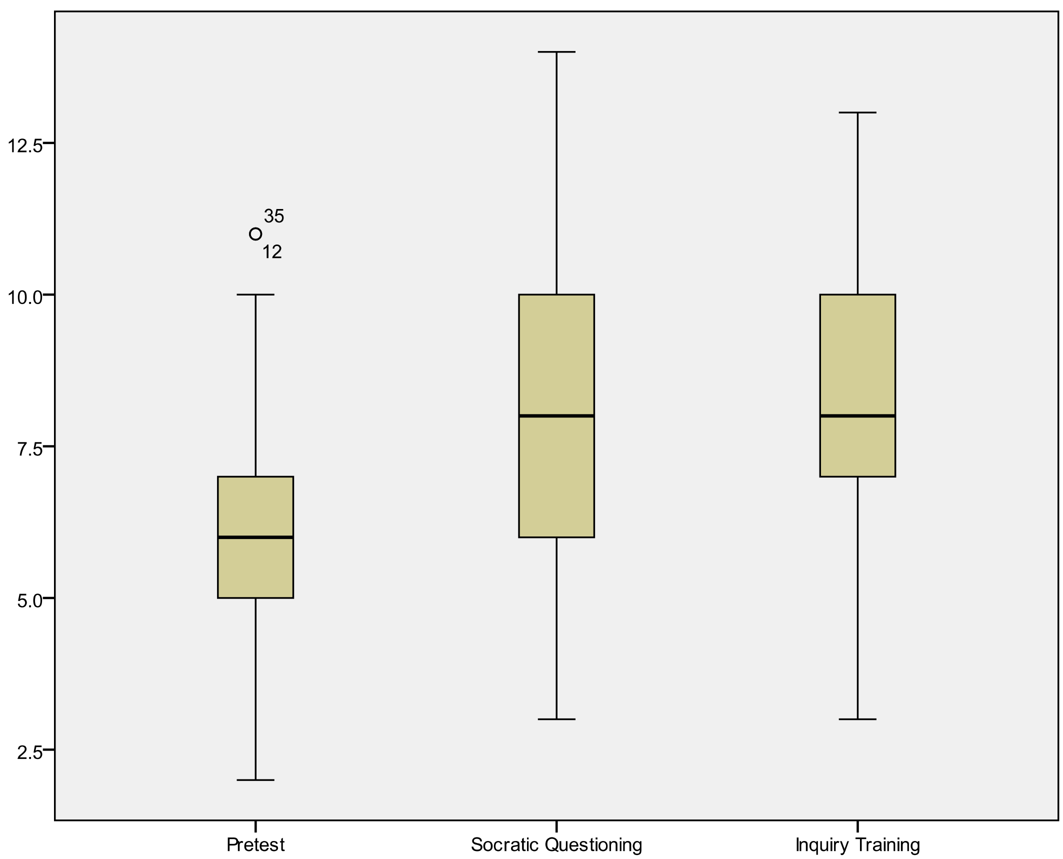
Descriptive Statistics

	N	Mean
Pretest	44	5.93
Socratic Questioning	44	7.95
Inquiry Training	44	8.09
Valid N (listwise)	44	

Explore

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Pretest	44	100.0%	0	.0%	44	100.0%
Socratic Questioning	44	100.0%	0	.0%	44	100.0%
Inquiry Training	44	100.0%	0	.0%	44	100.0%



References

Anderson, Lorin W.. "Bloom, B. S. (1913–1999)." Encyclopedia of Education. 2003. Retrieved April 04, 2010 from Encyclopedia.com: http://www.encyclopedia.com/doc/1G2-3403200074.html

Braundy, M. (2004). *Dewey's technological literacy: past, present, and future.* Journal of Industrial Teacher Education, 41(2), Retrieved from ERIC database.

Foundation for Critical Thinking. (2009). Criticalthinking.org -. Retrieved 04 April 2010, from The Critical Thinking community http://www.criticalthinking.org/index.cfm

Gravetter, F.J. & Wallnau, L.B. (2005). *Essentials of statistics for the behavioral sciences* (5th ed.). Belmont, CA: Wadsworth/Thomson Learning.

ISBE. (2010). Illinois standards achievement test (isat). Retrieved 06 March 2010, from Illinois State Board of Education http://www.isbe.state.il.us/assessment/isat.htm.

Northwest Evaluation Association. (2010). Nwea - reports site. Retrieved 06 March 2010, from NWEA Reports https://reports.nwea.org/ReportOrdering/ReportOrder.aspx.

Stillings, N. A., Ramirez, M. A., & Wenk, L. (1999). *Assessing critical thinking in a student-active science curriculum.* Paper presented at the meeting of the National Association of Research on Science Teaching, Boston, MA. http://scholar.google.com/scholar