ABSTRACT

STARLING, TINA TEDDER. The Effects of Self-Assessment on Achievement in an Algebra I Classroom. (Under the direction of Dr. Lee V. Stiff.)

Students in an Algebra I classroom were asked to self-assess their learning of concepts involving the real number system. A self-assessment tool, the Capacity Matrix, was designed to emulate Bloom's taxonomy and provide students with a means of reflecting on their understanding of the topics and concepts throughout the unit. A second group of Algebra I students received identical instruction throughout the unit on the real number system. They were not asked, however, to perform any formal self-assessment. A two-sample t-test showed that there was no significant difference between the achievement scores of the treatment group and the achievement scores of the control group. However, using Pearson Product Moment correlations, strong positive correlations between the self-assessment scores and the achievement scores were evident with the subgroups containing male students and with subgroups containing non-white students.

THE EFFECTS OF SELF-ASSESSMENT ON ACHIEVEMENT IN AN ALGEBRA I CLASSROOM

By

TINA TEDDER STARLING

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

MATHEMATICS EDUCATION

Raleigh

2005

APPROVED BY:

____________________________ ____________________________

____
Chair of Advisory Committee

BIOGRAPHY

Tina Julian Tedder Starling was born in Statesville, NC on August 17, 1977. She is the daughter of Randy and Sherry Tedder. Tina graduated from North Iredell High School in 1995, and was a recipient of the North Carolina Teaching Fellows Scholarship. She graduated from NC State University in May, 1999 with a Bachelor of Science degree in mathematics and in December, 1999 with a Bachelor of Science degree in mathematics education. In January, 2000, she was employed by Athens Drive High School in Raleigh, NC. After a year and a half of teaching, Starling decided to pursue a Master's degree part-time. For four and a half years, Starling continued teaching full time while taking graduate courses. Upon completion of her Master of Science degree, she plans to continue teaching mathematics and stay active in the mathematics education community.

ACKNOWLEDGEMENTS

I express my gratitude to my committee chair, Dr. Lee V. Stiff, for his guidance, advice, vision, and support. I also thank Dr. Richard Chandler and Dr. William Waters for serving on my advisory committee and the faculty and staff in Mathematics Education at NC State for their support through this academic journey.

Of course, none of this would be possible without the blessings from my Heavenly Father. I am personally grateful for loving, Christian parents who taught me about those blessings. They also taught me the importance of education - gently pushing me, praying for me, and encouraging me in my pursuit of an advanced degree. Finally, I thank my husband, Ray, for his prayers, patience, and love throughout graduate school. His genuine interest and incessant thoughtfulness helped me persevere. This thesis is dedicated to him.

TABLE OF CONTENTS

	Page
LIST OF FIGURES	v
INTRODUCTION	1
METHODOLOGY	
Purpose	7
Participants	7
Instrument	7
Procedure	9
RESULTS	14
DISCUSSION	19
CONCLUSION	23
BIBLIOGRAPHY	26

LIST OF FIGURES

Page

Figure 1: Capacity Matrix used by the Treatment Group 11

Figure 2: Self Assessment Scores by the Treatment Group on Fact Teams 12

Figure 3: Self-Assessment Averages of the Treatment Group on Fact Teams ... 14

Figure 4: Average IAAS and ICAS for Subgroups of the Treatment Group 17

Figure 5: Results of the Two-Sample T-test 18

INTRODUCTION

Often in education, the terms "evaluation" and "assessment" are used interchangeably. Evaluation has been defined by NCTM as "the process of determining the worth of, or assigning a value to, something on the basis of careful examination and judgements" (1995). Assessment, on the other hand, refers to collecting and analyzing data before making judgments. It is "the process of gathering evidence about a student's knowledge of, ability to use, and disposition toward, mathematics and making inferences from that evidence for a variety of purposes" (NCTM, 1995). This study demonstrates the effects of self-assessment by Algebra I students on their achievement as determined by an evaluation at the end of the study unit.

Existing research claims that self-assessment, sometimes referred to as self-regulation, positively influences student achievement in the classroom. Self-awareness of performance is a part of self-regulation (Alderman, 1999) and students who are able to self-regulate and monitor progress are more likely to be successful. "What is clear is that successful learners are highly self-regulated. Promoting self-regulation should be a major goal of instruction in any classroom setting" (Bruning, Ronning, & Schraw, 1999). Students should be encouraged to accurately rate themselves as learners in order to enhance their success as students. "Students who cannot monitor their own learning are at a great disadvantage. They are likely to continue to make the same mistakes. Self-monitoring serves as a tool for self-improvement and enhances learning" (Alderman, 1999). Teresa Garcia also discovered, in her 1992 study, that "metacognitive self-regulatory strategies were consistently positively related to critical thinking." Since both critical thinking skills and mathematical reasoning greatly influence achievement in mathematics, promoting students' ability to assess their mathematical performance should be important to overall achievement.

While self-assessment seems to augment understanding in most cases, one study suggests that such an activity may bring forward emotions which hinder the learning process. "For some individuals [self-assessment] implies greater concern about doing well and higher positive self-evaluation. For others, who are threatened by failure, heightened concern about self-evaluation is associated with an increase in reported anxiety implying the greater weighting of negative feelings" (Atkinkson & Raynor, 1974). But according to a more recent study, correlations between anxiety and academic achievement were often found to be weak. The study reports that these correlations are rarely able to explain more than 10% of the achievement scores (Geotz, Pekrun, Perry, & Titz, 2002). Hence, as the more recent study suggests, in mathematics, heightened anxiety does not always work against students.

There are many factors which motivate students. Students are affected by personal influences such as their own goals and academic abilities. They are also affected by situational influences such as classroom environment, and teacher and parental support. Both personal and influential elements give students cues about how well they are learning. Educators know from experience that the motivation level of a student increases when a student perceives that he or she is performing well. Researchers also agree that positive academic emotions promote academic motivation. In their study, Geotz, Pekrun, Perry, & Titz used two scales developed for the purpose of measuring students' perceptions (2002). The results illustrate significantly positive correlations between positive academic emotions and perceived self-regulation as well as predicted high achievement. Therefore, for the student performing well already, self-assessment should only increase his or her motivation and have no adverse affect on his or her achievement. Self-assessment also has been found to be helpful for students performing at lower levels. Research suggests that, with regard to self-assessment, "lack of success or slow progress will not necessarily lower self-efficacy and motivation if learners believe they can perform better by adjusting their approach" (Eccles & Wigfield, 2002). Thus, a self-assessment tool should maintain or enhance the learning of all students so long as it is used correctly.

Students must be given the opportunity to take ownership of their learning. This idea stems from the works of philosopher W. Edwards Deming and his student, David Langford. Langford suggests that "when schools underestimate the ability [of students] to identify [learning] needs…, the result is the teacher-directed, meaningless process that is endemic to many classrooms" (1995). The self-assessment tool used in this study, called the Capacity Matrix, was adapted from the competency matrix developed by Langford. It "provides a way for students and teachers to have ready access to students' learning process. It also encourages students to learn from failures" (Cleary & Langford, 1995). Rather than students feeling like they have mastered a concept or not, students should use the matrix to determine how their understanding of the content has developed.

Langford's competency matrix uses the six levels of understanding developed by Benjamin Bloom in the 1950s as the basis of his model. While Bloom's taxonomy is well-known among preservice and inservice teachers, researcher David Sousa suggests "that the taxonomy's value as a model for moving all students to higher levels of thinking has barely been explored" (2001). The six levels of Bloom's taxonomy include Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation. Bloom's taxonomy "remains one of the most useful tools for moving students, especially slower learners, to higher levels of thinking" (Sousa, 2001).

Students performing on the Knowledge level of Bloom's taxonomy have used memorization techniques and strategies and are able to recall facts. "It represents the lowest level of learning in the cognitive domain because there is no presumption that the learner understands what is being recalled" (Sousa, 2001). On the Comprehension level, students understand the information they learned in the Knowledge level. They are able to generalize from the situation in which the rule, algorithm, concept, principle or procedure is learned to a new situation. The use of algorithms in computation, production of new examples, translating verbal expressions into algebraic expressions and algebraic expressions into verbal expressions are all examples of Comprehension. The third level of Bloom's taxonomy, Application, involves students possessing the ability to use learned material in new situations with little or no direction from the teacher. Students on this level have the ability to perform straightforward combinations of previously learned principles put together in a new way. Analysis is the ability to partition material into components. At this level, a student will be able to break the new problem down into subproblems that he or she already knows how to do. Students performing on the Synthesis level should be able to create something unique to themselves that expresses understanding. They are able to make generalizations and put information together in a way that is different from that directly suggested by the statement of the problem. They "use divergent thinking to get an *Aha!* Experience" (Sousa, 2001). Finally, students at the Evaluation level can make a judgment about similar information and support this judgment using skills and information learned previously (Brown, 2004). Specifically with respect to mathematics education, students on the Evaluation level should be able to make judgments about the value of the mathematics and its uses.

Bloom's taxonomy provides a concrete structure that assists teachers in helping students develop their critical thinking skills. "Perhaps the most important aspect of Bloom's taxonomy is that it teaches thinkers to be critical of their own thinking. It reassures awareness and assessment of the thinking process itself, creating metacognition. Without this awareness and self-assessment, students are not critical thinkers" (Brown, 2004).

There are limitations, however, to Bloom's taxonomy. The structure of the hierarchy is inflexible. Brown argues, in her research, that very few learners start at the bottom of the taxonomy and progress forward one level at a time. Students may come to class with a wealth of knowledge about the subject already. "It is entirely possible that a learner may come to a situation with a pre-existing analysis or evaluation and then pursue the facts to back up the assumptions" (Brown, 2004). Nonetheless, Brown convincingly argues that the taxonomy does produce positive results. Her research shows that "utilization of this taxonomy should result in a classroom full of thinkers that are capable of establishing clarity and accuracy, of assessing relevance, and demonstrating the ability to think with depth, reach, and logic: skills that are fundamental to critical thinking" (2004).

Although there has been little to no research about the specific relationship between self-assessment and achievement in an Algebra I classroom, there is enough existing evidence that self-assessment can influence or motivate learning. In turn, motivation affects achievement. Since students' evaluation of themselves has more of an impact on improvement than evaluations made only by teachers (Cleary & Langford, 1995), and Bloom's taxonomy is a useful tool for moving students to higher levels of thinking, Langford's competency matrix is fitting for the mathematics classroom used in this study. The competency matrix allows students to evaluate the "kind of learning that students have at a given point" and that the completed matrix reflects "knowledge, understanding, and thinking skills such as application and analysis" (1995). The adapted tool used in this investigation, called the Capacity Matrix, employed a variation of Bloom's taxonomy of the levels of understanding. This self-assessment instrument is described at length in the methodology.

METHODOLOGY

Purpose

The purpose of this study was to determine if self-assessment by students in the treatment group resulted in higher achievement scores than the achievement scores for students in the control group. In turn, the treatment group was examined to decide if there was a correlation between a student's self-assessment and his or her level of achievement. One variable involved in the study was students' self-assessment of their understanding of operations with real numbers. The self-assessment instrument that was used is called a Capacity Matrix. Scores on the end of unit test were used to measure student achievement.

The null hypothesis tested in this study was:

Null Hypothesis: There is no significant relationship between the achievement scores of the control group and of the treatment group.

Participants

Two groups of high school students enrolled in Algebra I courses were used in this study. The control group (2nd period) consisted of 20 students – 11 females and 9 males. The treatment group (4th period) consisted of 29 students – 18 females and 11 males. Both groups included students from various age, socioeconomic, and racial groups. Additionally, based on the class test averages, both groups performed at the same academic level during the unit of instruction prior to the researched unit.

Instrument

The self-assessment instrument that was created, the Capacity Matrix, was modeled after that of David Langford. According to Langford, when all of the outcomes of a unit are listed, and students' progress is charted with respect to these outcomes rather than to some arbitrary standard for the unit, students and teachers alike have a clear understanding of student progress (Langford and Cleary, 1995). By using the matrix as a means of self-assessment, the student becomes "increasingly aware that learning is not binomial ("I either know it or I don't know it"), but progressive and dynamic" (1995).

In this study, the topics and concepts of the unit, as well as four levels of understanding, were used to design the Capacity Matrix. The topics and concepts of the unit were as follows: Problem-Solving Strategies, Commutative Properties, Fact Teams, Integers, Addition, Subtraction, Multiplication, and Division of Integers, Powers, Order of Operations, Associative Properties, and the Distributive Property. These topics and concepts, as outlined by Algebra I: A Process Approach, were not taught simultaneously. Instead, using this text, students had the opportunity to learn the topics and concepts over time by working questions of the consecutive problem sets within the unit. The following table illustrates the pacing guide for the topics and concepts emphasized throughout the unit:

Topic/Concept	**Day of the Unit**
Problem-Solving Strategies	1
Commutative Properties	1,2,3,4,5,6,7,12
Fact Team	1,2,3,4,5,6,7,12
Integers	2,3,4,5,6,12
Addition	3,4,5,6,7,8,9,10,11,12
Subtraction	4,5,6,7,8,9,10,11,12
Multiplication	6,7,8,9,10,11,12
Division	7,8,9,10,11,12
Powers	8,9,10,11,12
Order of Operations	9,10,11,12
Associative Properties	1,10,11,12
Distributive Properties	10,11,12

The four levels of understanding were designed to emulate Bloom's taxonomy. Although Bloom used six levels of understanding, Langford suggested that the three highest levels (analysis, synthesis, and evaluation) could be combined. The matrix reports "reflected knowledge, understanding, and thinking skills such as application and analysis" (Langford and Cleary, 1995). The Capacity Matrix used in this study uses the following four levels of understanding: I've Heard Of It, I Get It, I Can Apply It, and I Can Explain It. These four levels of understanding are intended to be equivalent Bloom's levels of Knowledge, Comprehension, Application and the highest three levels combined.

Procedure

Two classes of Algebra I students were used. Both groups were taught the same lessons by the same teacher. The two groups were assigned the same problems for homework and were provided with the same experiences during class. They were also given identical means of formal evaluation at the end of the unit. The treatment group was asked to do one additional task in class on six different days of instruction – self assess their own learning.

Students in the treatment group assessed their knowledge of each topic/concept by shading the corresponding spaces created in the 12 x 4 matrix (see p.11). For each of the topics and concepts previously listed, students assessed whether they had "heard of it," whether they "got it," whether they could "apply it," or whether they could "explain it." Students also justified several of their responses through journals, including examples of their knowledge. Langford, in his research, asked students to demonstrate their knowledge "so that the matrix reflects a realistic assessment of progress" (Langford and Cleary, 1995). The treatment group was asked to complete the instrument throughout the unit.

Days of self-assessment were chosen at random. Students may not have taken the assessment seriously if it were an every-day occurrence, and permitting time between assessments forced students to reflect on several days of instruction and learning. One self-assessment was administered on the first day, before any instruction for the unit began. Students in the treatment group were given very few instructions on the first day except to complete it without discussion with their peers and that the assessment would be kept confidential with the teacher. There was not a corresponding journal with the first assessment. Before the second self-assessment on fourth day, students were further instructed to be honest about their abilities and were asked to also complete the bottom portion of the self-assessment by writing a short journal about their responses. The teacher instructed students to choose the level of understanding that they shaded the most and explain what they could do to verify they were at that level. Subsequent self-assessments were assigned at random and occurred on the sixth, ninth, tenth, and twelfth days of the unit. Journals were completed on the fourth, sixth, ninth, and twelfth days.

Figure 1: Capacity Matrix used by the Treatment Group

Algebra I Chapter 2: The Real Numbers
Capacity Matrix[1]

Name:________________
Date:_________________

Chapter 2 Topics/Concepts	I've Heard of It	I Get It	I Can Apply	I Can Explain It
Problem-Solving Strategies				
Commutative Properties				
Fact Teams				
Integers				
Addition				
Subtraction				
Multiplication				
Division				
Powers				
Order of Operations				
Associative Properties				
Distributive Property				

Today, what did you pick the most?______________________________
What can you do that justifies why you chose that?______________________

[1] Adapted from David P. Langford and Barbara A. Cleary, Ph.D. (1995).

Spreadsheets were created for the topics/concepts in the unit. Figure two below shows students' self-assessment scores for Fact Teams, the third topic and concept from the unit, on the first, fourth, sixth, ninth, tenth, and twelfth days of the unit. A point value was assigned for each of the levels of knowledge: "I've Heard of It" was worth one point, "I Get It" was worth two points, "I Can Apply" was worth three points, and "I Can Explain It" was worth four points. Over the course of the unit, records were kept on students' self-assessment of each of the topics/concepts.

Figure 2: Self Assessment Scores by the Treatment Group on Fact Teams

Student	***First***	***Fourth***	***Sixth***	***Ninth***	***Tenth***	***Twelfth***
1	3	3	3	3	4	3
2	0	2	1	1	4	4
3	4	4	4	4	4	4
4	1	4	4	4	4	4
4	4	2			4	4
5	4	4		4		4
6	3	3	4	4	4	4
7	4	4	4	4	4	4
8	2	4	4	4	4	4
9	4	4	4	4	4	4
10	4	4	4	4	4	4
11	4	4	4	4		4
12	2	4	4	4	4	4
13	4	3	4	4	4	4
14	3	4	4	4	4	4
15	3	3	2	3	3	3
16	3	0	4	4		4
17	2	2	4	3	3	2
18	4	4	4	4	4	4
19	3	3	3	3	4	4
20	4	4	4	4	4	4
21	3.5	4	4	4	4	4
22	2	2	2	2	4	4
23	1	3	0	0	1	3
24	4	4	4	4	4	4
25	4	4	4	4	4	4
26	2	2	2	3	3	
27		1	2	4	4	
28	4	4	3	4	4	4

The average self-assessment scores for each individual student as well as for the class were calculated for each of the twelve topics and concepts of the unit. When the data collection was complete, a Pearson product moment correlation was used to analyze the relationship between the self-assessment scores and the test scores of the students in the treatment group. The main objective, however, was to determine if there was a significant difference in achievement between the control and treatment groups. Academic achievement was measured with a unit test. A standard two-sample t-test was used to test the null hypothesis that there is no significant relationship between the achievement scores of the control group and of the treatment group.

Unit test scores were collected from the control group, and unit test scores and individual concept assessment scores were collected from the treatment group. Once all data were collected, the results were analyzed in two ways. First, an analysis was done of the treatment group data alone to look for relationships and patterns. Second, an analysis using unit test grades from both the control group and treatment group was completed.

RESULTS

The average self-assessment score for each topic and concept for each individual student was calculated (individual concept assessment score, ICAS). Figure three below shows the data displayed previously in figure two with additional results. The right column shows the ICAS for Fact Teams. The same averages were calculated for the other eleven topics and concepts from the unit.

Figure 3: Self-Assessment Averages of the Treatment Group on Fact Teams

Student	**First**	**Fourth**	**Sixth**	**Ninth**	**Tenth**	**Twelfth**	**Average**
1	3	3	3	3	4	3	3.1666667
2	0	2	1	1	4	4	2
3	4	4	4	4	4	4	4
4	1	4	4	4	4	4	3.5
4	4	2			4	4	3.5
5	4	4		4		4	4
6	3	3	4	4	4	4	3.6666667
7	4	4	4	4	4	4	4
8	2	4	4	4	4	4	3.6666667
9	4	4	4	4	4	4	4
10	4	4	4	4	4	4	4
11	4	4	4	4		4	4
12	2	4	4	4	4	4	3.6666667
13	4	3	4	4	4	4	3.8333333
14	3	4	4	4	4	4	3.8333333
15	3	3	2	3	3	3	2.8333333
16	3	0	4	4		4	3
17	2	2	4	3	3	2	2.6666667
18	4	4	4	4	4	4	4
19	3	3	3	3	4	4	3.3333333
20	4	4	4	4	4	4	4
21	3.5	4	4	4	4	4	3.9166667
22	2	2	2	2	4	4	2.6666667
23	1	3	0	0	1	3	1.3333333
24	4	4	4	4	4	4	4
25	4	4	4	4	4	4	4
26	2	2	2	3	3		2.4
27		1	2	4	4		2.75
28	4	4	3	4	4	4	3.833333
AVERAGES	3.053571	3.206897	3.3333	3.5	3.769231	3.814815	3.433333

The mean of those averages was also found (individual average assessment score, IAAS). For the treatment group, a Pearson product moment correlation was run to assess the relationship between the individual average assessment score, IAAS, and the unit test grade. This analysis revealed some relationship between a student's individual average assessment score, IAAS, and his or her unit test grade ($r = .747986$). Subsequent Pearson product moment correlations were run to assess the relationship between the individual concept assessment scores, ICAS, and the unit test grade. The following chart gives the topics and concepts from the unit and the corresponding Pearson product moment correlations.

Problem-Solving Strategies	r = .518218
Commutative Properties	r = .543547
Fact Teams	r = .518218
Integers	r = .601745
Addition	r = .640699
Subtraction	r = .677636
Multiplication	r = .674239
Division	r = .632364
Order of Operations	r = .595347
Powers	r = .418823
Associative Properties	r = .413632
Distributive Property	r = .478405

The above correlations reveal no significant relationships between a student's ICAS and his or her unit test grade.

The treatment group was then partitioned into various subgroups. Pearson product moment correlations were used once again to assess the relationship between individual concept and average assessment scores and the unit test grades. The results are listed in Figure 3. The most promising r-values are those of the male students in the treatment group. Five out of the twelve individual correlations were greater than $r = .7$ with the correlation associated with order of operations as high as .80663. The overall IAAS for the boys was .77188. Other subgroups, such as the non-white students, and the white and non-white boys had similar and often stronger positive correlations. Six out of the twelve individual correlations for non-white boys were greater than $r = .7$, with four of those correlations also being greater than $r = .8$. These values support the claim that the self-assessment exercises positively influenced this subgroup of students.

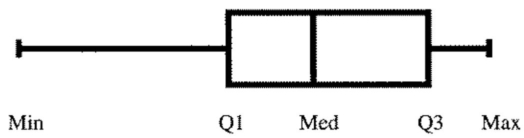
Figure 4: Average IAAS and ICAS for Subgroups of the Treatment Group

Subgroups of the Treatment Group With Corresponding Pearson Product Moment Correlations

	Girls	Boys	Non-Whites	Whites	Non-White Girls	Non-White Boys	White Girls	White Boys
Number in Group	18	11	13	16	7	6	11	5
IAAS	.70075	.77188	.67865	.70439	.59195	.74353	.23569	.83063
Problem-Solving	.58376	.42286	.36074	.13001	.48298	.31651	.34902	-.3367
Commutativity	.34385	.68148	.46843	.29417	.25103	.59319	.00231	.83728
Fact Teams	.58376	.42286	.36075	.13001	.48298	.31651	.34902	-.3367
Integers	.30015	.73392	.51478	.58419	.04449	.74576	.46095	.55997
Addition	.42277	.75431	.70011	.18804	.38719	.8368	.43505	*
Subtraction	.55779	.75431	.73109	*	.48364	.8368	*	*
Multiplication	.55779	.73893	.71544	*	.48634	.81659	*	*
Division	.59071	.65361	.63883	.12664	.49183	.70304	.33102	*
Order of Operations	.20351	.80663	.57409	.58867	.12207	.82211	.43558	.72542
Powers	.24953	.53062	.27611	.47024	-.0437	.43836	.49569	.78106
Associativity	.11686	.64294	.41844	.2763	.20921	.56331	-.0915	.76676
Distributive Property	.27813	.59569	.37033	.43136	.27131	.50288	.1212	.7765

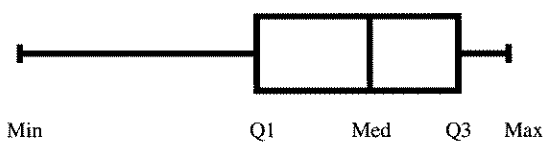
* Students in this subgroup marked "I Can Explain It" on each of the six self-assessments. Therefore they all received an IAAS of 4. Hence, a Pearson product moment correlation is not applicable in such cases.

Achievement was measured by scores on a unit test. The control group consisted of 20 students with a mean unit test score of 89.55. The standard deviation of the group was 7.35903. Other one-variable statistics for the control group were as follows:



Minimum:	73
First Quartile:	85
Median:	90
Third Quartile:	96.5
Maximum:	100

The treatment group consisted of 29 students with a mean unit test score of 90.24. The standard deviation of the group was 6.56644. Other one-variable statistics for the treatment group were as follows:



Minimum:	71
First Quartile:	84.5
Median:	91
Third Quartile:	96
Maximum:	99

In order to test the null hypothesis, namely that there was no significant difference between the achievement scores of the control group and of the treatment group, a two-sample t-test was used. The results from that test showed $t = .33758$, and $p = .73755$ with approximately 37.85046 degrees of freedom. Therefore, since p is not less than $\alpha = .05$, there was failure to reject the null hypothesis. The results are summarized in the figure below.

Figure 5: Results of the Two-Sample T-test

Group	Number of Students	Mean	Standard Deviation
Control	20	89.55	7.35903
Treatment	29	90.24	6.56644

$t = .33758$, $df = 37.85046$, $p = .73755$

No significant difference between the means of the control and treatment groups.

DISCUSSION

The statistical results, from the two-sample t-test, show that there was no significant difference between the achievement scores of the control group and the achievement scores of the treatment group. That is to say, there was no significant difference between the unit test scores of the students who were not exposed to any formal model of self-assessment and the unit test scores of the students who were asked to complete the Capacity Matrix six times throughout the unit. To understand this fully, the treatment group was examined in greater detail.

When the Pearson product moment correlation was run on the IAAS and the unit test scores of those students in the treatment group, there was some relationship ($r = .747986$). However, there were no significant relationships found between any of the ICAS or the unit test scores in treatment group students. The highest Pearson product moment correlation was $r = .677636$, the correlation between the unit test scores and the self-assessment scores on the concept of subtraction. Some correlations were as low as $r = .413632$, the correlation found when comparing the unit test scores and the associative properties.

Having found no significant relationship among the ICAS or the unit test scores, the treatment group was examined even more closely. The second row in Figure 3 shows the subgroups of the treatment group's corresponding Pearson product moment correlations of the unit test scores and the IAAS. The mean correlation of all subgroups was $r = .657184$. However, the values ranged from $r = .23569$, the correlation associated with the white girls subgroup, to $r = .83063$, the correlation associated with the white boys subgroup. Interestingly, there was a stronger correlation between the self-assessment and the unit test of the boys than the girls in all but two subgroups of the topics and concepts: Problem-Solving and Fact Teams. Thirteen students in the treatment group reported that the self-assessment exercises actually helped them during the unit. One male student stated, "it helped me because it made me think about what I know and don't know, so I can work on things." Another male student said, "it made me be honest with myself. Then I knew what to study harder on." Perhaps the self-assessments in this unit of study provided the male and minority students the opportunity to reflect on their learning – something that may not usually be encouraged.

Another interesting finding was that there was generally a stronger correlation between the self-assessment scores and the unit test scores of the non-white students than the white students in all subgroups. The only exceptions were with the concepts of integers, powers, and the associative properties where the Pearson product moment correlations for the white students were slightly greater than the corresponding correlations for the non-white students.

The subgroup consisting of only white girls appears to be the one that most strongly affects the overall significance. The eleven girls in this subgroup, on average, selected "I Can Explain It" in over eight topics and concepts. So, more than 67% of students in that subgroup were unable to notice any growth in their understanding of those concepts. This high percentage is in line with the unit test score average of those eleven students. Their mean test score was over 94.182. Without these students in the overall study, however, there is actually an average decrease of .5 from the correlations of the entire treatment group.

The measures of central tendency show a small improvement in average unit test scores of the treatment group over the control group of students. The means and medians for the treatment group were slightly higher than that of the control group. Even the mode for the treatment group was 91 while the mode for the control group was 90. Nonetheless, because the five critical values for creating a statistical box-plot are so similar, it is difficult to say that the self-assessment made a positive impact on the students' academic achievement in the unit. For the same reason, however, one may not suggest that the self-assessment in any way hindered the academic performance of the students. Many self-assessment averages for unit topics and concepts increased throughout the unit. The self-assessment scores for Problem-Solving, Fact Teams, and Integers steadily increased from the first day of the unit to the twelfth day of the unit. The probable cause for this was the fact that succeeding topics and concepts depended on the understanding of these three topics and concepts. Other self-assessment scores illustrated slight increases and decreases during the unit, but showed overall higher scores on the twelfth day. Order of Operations, Powers, Associative Properties, and the Distributive Property were topics with the highest self-assessment scores occurring on the last day. More than likely, students thought they knew these topics better than they did, were exposed to difficult problems on days nine and ten of the unit and, with more practice on days eleven and twelve, felt more comfortable with the topics at the end of the unit.

Still other topics and concepts possessed self-assessment scores that were continually and inconsistently up and down. For example, the self-assessment scores for Division started high, decreased a little, peaked on the ninth day, decreased a little again and then ended with the second highest average score. Also, the Multiplication, Addition and Subtraction scores started on day one with an average self-assessment score of 4. In other words, everyone could "Explain It". As the problem sets in the unit continued, however, their level of understanding shifted with some scores of two and three. On the first day, students probably thought of the multiplication, addition and subtraction problems they had done for several years. Once negative real numbers were introduced, they were not as confident in their understanding. The self-assessment scores for the Commutative Properties appear somewhat as a normal curve. The students' perception of understanding was highest on day nine of the unit with lower scores before and after. Their self-assessment scores on the properties started low, peaked on the ninth day, dropped on the tenth day, and increased to the second highest score on the twelfth day of the unit. Perhaps this is because the students were unfamiliar with the Commutative Properties at the beginning of the unit and the textbook did not emphasize the commutative properties after day seven until the twelfth day. Regardless of the topic or concept, it seemed evident that the students' self-assessment of their understanding was in line with the amount of emphasis placed on the topic or concept by the authors of the text. Still, there were some students who perceived themselves to be ones who could "Explain It" from day one and did not perform as well on the test as students who were better able to accurately reflect their own understanding.

CONCLUSION

Research has shown that self-regulation, or self-assessment, is important and necessary for students at all levels. Self-assessment exercises encourage students to be reflective of their own learning and this, in turn, accelerates the critical thinking skills particularly useful in an Algebra I class setting. Students who can accurately perceive their own level of understanding are at a great advantage. They recognize where they stand with respect to their learning and where they need to go to have a deeper understanding. Bloom's taxonomy, while introduced in the 1950s, continues to be the choice model for moving students to higher levels of thinking.

The Capacity Matrix used in this study was a self-assessment tool which emulated Bloom's taxonomy. The assumption, that self-assessment experiences would result in higher academic achievement as determined by a unit test, was tested. The treatment group of students that used self-assessment throughout the unit did, in fact, have higher measures of central tendency than the control group. The difference between those measures of the treatment and control groups, however, was not significant.

Nevertheless, the self-assessments performed throughout the unit seemed to suggest some positive impact on the achievement of students in certain subgroups. Although all correlations were weak, there was an overall stronger positive correlation between the self-assessment and unit test grades for the male and minority students. Further study in the area of self-assessment versus achievement among male and minority students is necessary to determine the validity of the argument.

The correlations between the self-assessment and achievement may be strengthened by collecting more data over several months and several units of study. Moreover, by gathering additional background information the researcher may be able to make more inferences about the implications of self-assessment on learners in many subgroups. More specific instructions and discussion at the beginning regarding the topics and concepts of the unit and the levels of learning used in the Capacity Matrix would also, perhaps, result in more accurate measurements of self-assessment. Using Addition of Integers, Subtraction of Integers, Multiplication of Integers, and Division of Integers instead of keeping integers and the operations with integers separate made have made it easier for the students to honestly reflect on their abilities. Practice with the self-assessment matrix and the meaning with the levels of understanding, allowing students to become comfortable with what "I Can Apply It" and "I Can Explain It" really means would be useful in future studies. To decide whether self-assessment positively impacts academic achievement, students need the ability to self-regulate and accurately reflect on their own level of understanding. More activities and directions regarding self-regulation prior to the self-assessment with the Capacity Matrix would have been helpful in this study. Still, the use of self-assessment tools, such as the Capacity Matrix, certainly did not hinder the academic performance of any student. The achievement scores of the students were consistent with or better than their respective scores prior to this unit. Students also reflected on the self-assessment process and none commented that it impeded their academic progress through the unit.

Academic success for all students is the goal in education. Outside of the teacher's control are many factors that affect how students feel about their understanding of mathematical concepts and ultimately their academic performance in mathematics. If students can learn to accurately assess their own learning and communicate that assessment to the teacher, imagine how the teacher can then assess his or her own methods of instruction to better meet the needs of students. The notion of encouraging the use of formal self-assessment in any classroom, particularly the mathematics classroom, is a fairly new proposal. Further studies should be considered regarding the correlation between self-assessment and academic achievement in order for the proposal of classroom self-assessment to become a popular practice.

BIBLIOGRAPHY

Alderman, M. K. (1999). Motivation for Achievement: Possibilities for Teaching and Learning. Mahwah: Lawrence Erlbaum Associates.

Atkinson, J. W. and Raynor, J. O. (1974). Motivation and Achievement. Washington, D.C.: V. H. Winston & Sons.

Brown, T. (2004). Bloom's taxonomy: Bloom's taxonomy and critical thinking. In J. L. Kincheloe & D. Weil (Eds.). Critical thinking and learning: An encyclopedia for parents and teachers (pp. 77-82). Westport, CT: Greenwood Press.

Bruning, R. H., Ronning, R. R., & Schraw, G. J. (1999). Cognitive Psychology and Instruction (3rd ed.). New Jersey: Prentice-Hall, Inc.

Cleary, B. and Langford, D. (1995). Orchestrating Learning with Quality. Milwaukee: ASQC Quality Press.

Dougherty, B.J., Matsumoto, A.N., Rachlin, S.L. & Wada, L.T. (1994). Algebra I: A Process Approach. Honolulu: University of Hawaii.

Eccles, J. S. and Wigfield, A. (Eds.). (2002). Development of Achievement Motivation. San Diego: Academic Press.

Goetz, T., Pekrun, R., Perry, R. & Titz, W. (2002). Academic emotions in students' self-regulatory learning and achievement: A program of qualitative and quantitative research. Educational Psychology, 37(2), 91-105.

National Council of Teachers of Mathematics (1995). *Assessment Standards for School Mathematics.* Reston, VA: National Council of Teachers of Mathematics.

Sousa, D. (2001). How the Brain Learns (2nd ed.). Thousand Oaks: Corwin Press, Inc.