The Effect of Formative Assessment on Achievement and Motivation

By Martin Ashdale

Submitted in Partial Fulfillment of the Requirements for the

Degree of Master of Education

May 2020

Graduate Programs in Education

Goucher College

Table of Contents

List of Tables

Abstract

Formative assessments are teaching strategies used to gauge student understanding and make in class adjustments to improve a lesson. Formative assessments can also improve student achievement and motivation. While there is enough evidence that formative assessment significantly improves these student measures, there is insufficient evidence on specific formative assessment tools and strategies that are commonly used in the classroom by teachers. This study seeks to better understand the effect of a specific formative assessment strategy, Progress Trackers, by comparing a control group that did not receive the Progress Tracker with a treatment group that did receive the formative assessment. The study showed that there was no significant difference between the treatment and control group based on the means of a pretest and posttest score. While not statistically significant, the treatment group did show a larger increase of students with at least a 60% improvement in achievement. The lack of statistical significance between the control and treatment group could be caused by the ineffectiveness of the formative assessment or the inability to exclude other variables in the classroom setting. This could include the use of other formative assessments used in both groups, content delivery and implementation of the formative assessment. Adjustments to the Progress Tracker itself, implementation and use with other demographics would be needed to accurately determine the effectiveness or lack thereof of the Progress Tracker.

CHAPTER I

INTRODUCTION

Overview

Teachers are tasked with many duties over the course of a school year. One of those many duties is to assess student understanding of content and skills that students are required to learn according to the state standards for their subject area. Traditional assessments (quizzes, test and exams) are often used as evidence that students have attained the required content and skills. However, these assessments do not keep track of student learning during the instruction leading up to those traditional assessments. Most traditional assessments do not provide teachers with data that they can use with their current students but will only be useful for making adjustments during the following year when that content is taught again.

Formative assessments allow teachers to implement strategies during class time prior to a traditional assessment being administered. These strategies give both teachers and students feedback about their learning and allow both to adjust instruction prior to a traditional assessment. Because of the additional data students and teachers are better equipped to understand the gap in learning and try new ways of learning prior to a high stakes traditional assessment where teachers often move on from previous topics and continue with new material.

Statement of Problem

It is known that formative assessments are an effective tool for improving student understanding, academic achievement and motivation but there is not as much research on specific strategies. Teachers often hear of the benefits of formative assessments but struggle to identify and implement specific strategies within the classroom. This study looks to identify a

specific formative assessment strategy and determine whether there is significant evidence that this strategy will help to improve student achievement.

Null Hypothesis

The use of a Progress Tracker formative assessment strategy during a unit of instruction will not significantly increase student achievement compared to a control group.

Operational Definitions

This section describes and defines the key variables and concepts in this study.

Progress Tracker

A Progress Tracker is a digital tool that is used to collect student data at the beginning, middle and end of an instructional unit. This tool utilizes Microsoft Forms to collect this data. The progress tracker is composed of two main parts. This includes a rating of all the daily objectives for that unit of material and the second part is five selected response questions. During the first part of the Progress Tracker students rate their understanding of each daily objective on a scale of one to five. During the second part of the Progress Tracker students answer five selected response questions that are aligned to the big ideas of the unit. This can include multiple choice and true or false type questions.

Student Achievement

Students will be given a pre-test prior to any instruction and a post-test at the conclusion of instruction. Student achievement is defined as a 60 percent increase in score on a traditional selected response assessment. The experimental group will receive the Progress Tracker formative assessment treatment and will be compared to a control group that does not receive the Progress Tracker formative assessment treatment. Statistical analysis will be completed on the pre-and post-test scores to determine if there is a significant difference.

Independent Variable (IV)

Whether students were using the Progress Tracker or not**.**

Dependent Variable (DV)

The score on the pre- and post-traditional assessments.

CHAPTER II

REVIEW OF THE LITERATURE

Overview

Formative assessments are a relatively new tool that teachers can implement in their classroom to keep track of student progress, help students become more involved in their own learning and improve academic achievement for all students. This literature review explores six areas of formative assessments and discusses our current understanding of the effect of formative assessments and avenues where we can further our understanding. These areas of exploration are a definition of formative assessment, rationale, strategies, academic achievement, student motivation and professional development. Section one will define and describe formative assessments, section two explains why teachers should implement formative assessment strategies in their classroom, section three identifies and describe specific formative assessment strategies that can be implemented in a variety of classrooms, section four describes how formative assessment can increase student academic achievement, section five explains how formative assessments can improve student motivation and section six describe how teacher professional development can have a large impact on effective implementation of formative assessment in the classroom.

Definition of Formative Assessment

Formative assessment is a tool used by teachers to collect data on student understanding and performance. This data is used to make in class adjustments to lessons, differentiate instruction and for use in planning subsequent lessons (Heritage, 2007). Formative assessment can be embedded into traditional classroom teaching techniques or can be novel ways for checking student understanding. The use of formative assessment differs from previous

pedagogical techniques in that the student, teacher and student peers share the responsibility of learning and collaborate to reach the desired goal.

According to Heritage (2007), there are four core elements to formative assessment. These include, identification of the gap, feedback, student involvement and learning progressions. Formative assessments can be used to identify what a student already knows how to do or understands. When the teacher knows what the student can do, they can then build an instructional plan that scaffolds new skills and content to get them where they need to be (Tomanek, Talanquer & Novodvorsky, 2008). The teacher will also provide constructive feedback, so the student knows what they need to work on and think about how they will adjust themselves to achieve the learning goal. If a student is an active participant in their learning, they will be more motivated to participate, think critically and evaluate their learning. There is a high amount of collaboration between the teacher and student in evaluating their own work and the work of peers. Finally, teachers and students need to identify learning progressions. State standards are often broad and do not include the steps needed to get students from where they are to where they need to be. The teacher should help the student identify short term goals that lead to the ultimate desired goal of meeting the state standard. This may include the use of big ideas, essential questions and daily objectives. The student and teacher keep track of where the student is on this continuum throughout the unit or chapter.

Rationale for the Use of Formative Assessments

The use of formative assessments is often critical to the evaluation of a teacher through the popular teacher evaluation model of the Danielson Framework. The Danielson Framework is used as evaluation system that is broken down in four main domains. Those domains are Planning, Classroom Environment, Instruction and Professional Responsibilities. Each domain

is broken down further into components. Each component focuses on one aspect pertaining to the larger domain. The components are then ranked based on descriptive criteria of highly effective teaching for that component. Components 1f: Designing Student Assessments and 3d: Using Assessment in Instruction are areas where teachers are evaluated on their ability to assess students formatively and using traditional assessment techniques. All teachers are expected to teach the entirety of the curriculum and have students perform well on high stakes standardized testing. It has been shown that the use of formative assessment improves student understanding, comprehension and higher order thinking skills over traditional summative assessments. The use of formative assessments in the classroom setting can also reduce test anxiety and improve student performance on traditional summative assessments (Clark 2011).

The use of formative assessment is relatively new to education with most of the literature beginning from the 1980's and continuing to the present. Because the identification and use of formative assessments is relatively new, there is not much understood about how and why teachers use formative assessments in the classroom and how formative assessments fit into the larger picture of education.

Tomanek, et al., (2008) conducted a study to identify how and why teachers use formative assessment in the classroom. Their study investigated how and why both pre-service and experienced educators would or would not use certain formative assessment strategies using probes to elicit responses from the participants. They found that the type of formative assessment used would be decided by three main factors, expectations, characteristics and teaching experience. They found that both pre-service and experienced teachers had preconceptions of the formative assessment tasks that did not match with the educational researchers that both designed and studied the significant effects of the formative assessment

strategies. Specifically, in experienced teachers, they found that the experienced teachers would often choose high quality formative assessment strategies based on their belief that their students would not be able to complete the task.

The remainder of this literature review seeks to develop a better understanding of the specific formative assessment strategies that lead to an improvement in student achievement and motivation. It also investigates effective professional development of educators to help close the gap between what the research says about the use of formative assessment and the actual implementation of formative assessment by teachers in the classroom.

Formative Assessment Strategies

There are countless formative assessment strategies that can be used in the classroom. Many of the formative assessment strategies are techniques that are created based on the principles described in the previous section of this review. It was difficult to find studies with rigorous statistics of specific strategies. Most of the strategies below are well developed and are based on providing constructive feedback, determining the gap in knowledge, improving motivation and student involvement but most of the data associated with these strategies is purely anecdotal (Cauley & McMillan, 2010). All these criteria are based on the peer reviewed research of the benefits of formative assessment, but the specific strategies discussed in the articles have not been as rigorously reviewed as the source material on which they are based. Most of the strategies described in peer-reviewed research do not explain how the formative strategy was done but only explain it in the context of the study. Much of the research is qualitative and relies heavily on gathering data through observation and interviews as opposed to more quantitative measures.

Many formative assessment strategies are a combination of the multiple core pieces of formative assessment and include student involvement and feedback as key features of their formative assessment strategies. In a study completed by Anderson and Palm (2017), they found that the use of Assessment for Learning (AfL), learning intentions (LI) and success criteria (SC) were widely used but not completed with enough fidelity to yield the statistically significant results of the research. These strategies elicit student involvement by allowing them to see where they are, where they need to go and how they will get there. AfL focuses on the use of questioning techniques to elicit more detailed responses that will help to achieve the LI and SC. While both students and teachers were found to be aware of the purpose and could adequately define these terms, they had a hard time creating them and putting them into practice effectively.

In another descriptive study by Crumrine & Demers (2007), they highlight anecdotal evidence of formative assessments used by their colleagues at their high school. They highlight that traditional/common pedagogical strategies like the use of index cards, personal white-boards, journaling, questioning and re-teaching can be examples of high quality of formative assessment as long as they are implemented properly. These anecdotes describe how teachers that have not used formative assessment were shocked to see how little their student retained after weeks of instruction but showed improvement using the strategies stated strategies. There are also anecdotal accounts from students where they found that learning was made easier and engaging using high quality questioning and finding more success using the formative assessments.

In another descriptive study of science formative assessments at the elementary level, Martinez-Gudapkkam, Mutch-Jones, & Hicks (2017) found that changing their traditional approaches like giving meaningless, vague feedback and marking directly on papers was not as

effective at providing constructive, positive feedback as some other formative assessment strategies. Students showed improved motivation and an increased understanding of the science phenomena taught using both combining diagrams with writing and post-it note feedback. Traditionally, diagrams and explanations have been separated but the teacher allowed students to include a mixture of diagrams, data and writing in one space. This allowed students to demonstrate their understanding in the best way they saw fit. Additionally, using post-it notes to leave feedback allowed them to be removed once corrections were made. It did not leave a lasting mark on the assignment and allowed students to easily navigate the areas that they needed to work on.

One exception in terms of rigorous qualitative data on the use a specific formative assessment strategy is the use of technology enhanced formative assessment. In studies completed by Hwang and Change (2011) and Sheard and Chambers (2014), it was shown that the use of formative assessment in technology lead to significant increases in achievement. Both studies had students use a technology-based learning platform to complete a task and answer questions. While the controls of the studies used a traditional approach where, if a student received a wrong answer they were given the correct answer right away, the experimental groups were only given hints. Additionally, they cycled the same or similar questions back into the queue of questions that would be asked by the students and would only stop being asked after they were answered correctly each time. The combination of repetition and the use of hints instead of answers led to a significant improvement on the post-test scores. Additionally, in surveys taken by the students, it found that it also significantly increased motivation.

The Affect of Formative Assessment on Student Achievement

There are many ways to define student achievement but the most common is improved numerical score on either a teacher-generated summative assessment, or a state mandated standardized assessment.

In a review of the literature Black & Wiliam (2010) looked at the ability of formative assessments to improve achievement of students. They looked at studies where students that received an "innovation" (formative assessment) had significantly increased achievement than similar students without the innovation. They noted that low performing students saw an even greater improvement in achievement than already high achieving students.

In a study by Hudesman, Crosby, Flugman, Issac, Everson, and Clay (2013) it was found that a comprehensive formative assessment and self-regulatory learning lead to significant achievement improvement pass rates of associate level mathematics course. In the study they used a prescribed treatment of specific feedback, quizzes where students received specific feedback about their progress on specific assessment questions and students take surveys about their metacognition. The focus of this student is to use the feedback portion of formative assessment and having students thinking about their learning. This showed a significant improvement in achievement for those students.

Many of these studies have been completed by the pioneers of formative assessment, Black & Wiliam (2010) from the United Kingdom as well as many others from the United States but this study looks at formative assessment outside of the west, in Turkey. A study by Ozan and Kincal (2018) found similar findings to the other studies. The formative assessment strategies of success criteria, deliberate feedback and questioning techniques were used to determine if they influenced student achievement. They also looked at self-regulatory practices (metacognition).

They found that there was significant improvement in student achievement, motivation and metacognition.

In a study by Carrillo-de-la-Peäna, Bailláes, Caseras, Martâinez, Ortet & Pâerez (2009) found that formative assessments showed a significant improvement in undergraduate students majoring in health sciences. The students received immediate and deliberate feedback on quizzes during their term. The students that received the immediate feedback performed better on subsequent quizzes and on the summative assessments at the end of the term.

There is a direct link between all of these studies. Each study shows that there is a significant relationship between achievement and formative assessment. There are no specifics on the full implementation of formative assessments that accounts for other factors outside of the formative assessment itself.

The Effect of Formative Assessments on Student Motivation

The effect of formative assessment and motivation is also closely related to the results of achievement. In many of the students already mentioned, there have been a facet of the studies that also looked at motivation. It was found in every case that motivation also improved along with achievement.

In a study completed by Faber, Luyten and Visscher (2017), the authors found that after using an adaptive digital formative assessment tool, students were more motivated and felt better about their learning compared to a control group. The digital formative assessment was designed so that the teacher could choose specific assignments associated with topics that were being learned in math class. The adaptive portion of the digital formative assessment tool was that the types of questions and problems that the students received were based on their responses to previous questions. The types of questions that the students had to complete were always just

out of their current understanding. They were in the "goldilocks" zone where it was not too hard and not too easy for them to complete. This was accompanied by instantaneous feedback that would give students hints and provide them with additional resources.

In a descriptive study by Nolen (2011), found that motivation and the way that teachers perceived the motivation for students had a direct impact on the engagement and success of students in the classroom. The focus of the study was to investigate the connection between teacher feedback and the motivation of students. Rigorous and constructive feedback is often not used in the classroom with students who generally struggle, like those in poverty or low achieving students. The rigorous critique of work is generally given to students that have seen some sort of success in the classroom already. This is to prevent the students from losing faith that they will ever have success in the classroom, so teachers will often assess them on low-level criteria like neatness or organization. The study found that the use of rigorous formative assessment feedback may be very beneficial to students in that they are often asked to think about their learning. They must continually analyze, access and reflect on their learning. With the proper support and guidance from their teacher, this can provide challenges but also success for the student. This in turn leads to an increase in motivation and engagement from the student. This creates a positive feedback loop that begets more success.

Professional Development of Planning and Implementation of Formative Assessment Strategies in the Classroom

In many of the studies referenced in this review, the discussion mentions that there is not much clarity or even confusion on the part of the both the teachers and students that perform formative assessment in the classroom. They may not have a coherent idea how to structure formative assessments and determine exactly what it should look like in their classroom. The

following studies look to statistically analyze the effects of a full professional development program from train-the-trainer sessions, to breakout training, implementation and reflection.

In a study by Andersson and Palm (2017), the authors found that the implementation of a well-developed and widely used formative assessment program based on the work of other researchers in the field of formative assessment had teachers that participated in the professional development significantly outgains teachers of the control group in student achievement. The professional development that was used started from the very beginning by starting small and working toward training the entire staff. Similarly, to other studies, the professional development program did not focus on a single strategy but taught teachers about the main tenants of formative assessment, namely, feedback, student involvement, peer-assisted feedback and self-regulatory actions. After full implementation, completion and accounting for all variables they found that achievement on a post-test by the students of teachers that completed the program significantly outperformed those that did not.

In a descriptive study by Ash and Levitt (2003), the researchers used case studies to outline the importance of professional development. All their case studies were anecdotal and relied solely on teacher accounts instead of statistics, but they detail the relationship between mentor and mentee teachers in terms of their ability to implement formative assessment strategies. Each case study outlines the skills needed for effective formative assessment that includes feedback, communication with students, adjusting and reflection. The work that the mentor and mentee teachers do together allows the experienced teacher to guide and demonstrate their understanding of formative assessment and lets the mentee teacher learn from previous mistakes. This study demonstrates multiple detailed accounts of the importance of not only large-

scale professional development but the often smaller, mentor/mentee relationship that all novice teachers need to gain the skills to educate their students.

CHAPTER III

METHODS

Design

This study was conducted based on a quasi-experimental design utilizing a pre-test and post-test. The experimental group received the Progress Tracker formative assessment treatment during the duration of the instructional unit while the control group did not receive the Progress Tracker formative assessment. Prior to any instruction, both groups received a pre-test and a post-test after conclusion of the instructional unit. Pre-and post-test scores were compared at the conclusion of the instructional unit. The *t*-test statistical tool was used to determine if there was a difference in the population means of the treatment and control groups on the pre-test and on the post-test. To determine statistical significance, the *t*-test result was compared at the customary five percent alpha level. Descriptively the mean improvement from pre-to-post for the treatment and control samples was compared to a 60 percentage-point improvement.

Participants

The participants in the study were from ninth grade academic biology classes taught in a mixed urban/suburban high school. Approximately 66% of the school population was economically disadvantaged, 2% were English language learners, 22% were special education students, 52% were male and 48% were female.

For the study there are two groups of participants, a control group and an experimental group. The experimental group and control group contained combinations of classes that were co-taught, partially co-taught, and taught by one educator. The groupings of classes were random but deliberate in trying to keep similar class types spread as evenly as possible between

the control and experimental groups. All classes were taught by the same general educator, and co-taught classes, by the same special educator.

The control group contained three academic biology classes, including two classes that were co-taught 100% of the time and a class that was co-taught 50% of the time. The control group's demographics featured a total of 58 participants: 69% male, 31% female, 32.8% black, 53.4% white, 6.9% multi-racial, 3.4% Asian or pacific islander, 3.4% Hispanic and 37.9% special education students.

The experimental group participants also consisted of three academic biology classes. In these classes, one class was co-taught 100% of the time, one class was taught by one teacher and one class was co-taught 50% of the time. The experimental group's demographics featured a total of 57 participants: 50.9% male, 49.1% female, 52.6% black, 31.6% white, 10.5% multi-racial, 1.8% Asian or pacific islander, 3.5% Hispanic and 24.6% special education students.

Instrument

The instrument used in this study was a traditional assessment in a pre-test and post-test format. Both the pre-test and post-test consisted of selected response questions in a multiple-choice format. The content, style and number of questions were the same on both assessments; the questions tested parallel content on both assessments. The assessment was teacher-generated as most assessments, given throughout the school year, are generated by the teacher. Due to teacher generation of the assessment, there was no statistical evidence of validity, reliability or critiques of the assessment use. However, the teacher-researcher derived the items from released test items from the state assessment**.** The participants in the study were familiar with the style and format of the assessment, as they had received similarly styled assessments previously.

Procedure

Before any instruction of the unit material, the participants were asked to take the pre-test. The participants were allowed to write on the pre-test, but their answers were recorded on a bubble sheet that was graded using an electronic grading machine. The directions regarding the selection of only one response per question were written on the pre-test and communicated to the participants verbally. The participants were also made aware that the score on the pre-test would not affect their grade.

After the pre-test, the data was scored and a comparison of the mean between the experimental and control group was conducted to determine whether there was any significant difference between the experimental and control group before the start of the experimental treatment. Thus the non-significance of the pre-tests helped verify the efficacy of the sampling of classes to treatment or control.

The beginning Progress Tracker was then administered to the experimental group immediately after the pre-test. Midway through the instructional unit, the experimental group received the middle Progress Tracker and again the day before the post-test.

The same instructions were provided to both groups prior to the post-test except students were made aware that their score would affect their grade. After collection and scoring of the post-test, statistical analysis was conducted on the mean score of both the pre-test and post-test for the control and experimental groups.

The null hypothesis for the pretest was: There is no difference in the population mean pre-test score between the treatment and control group. The null hypothesis for the post-test was: There is no difference in the population mean post-test score between the treatment and control group.

CHAPTER IV

RESULTS

The purpose of this study was to determine if the use of the Progress Tracker formative assessment lead to a significant difference in achievement on a post-test over the pre-test scores. The results found that there was no statistical significance between the control and treatment groups after completing the pre-test. This suggests that both groups were even to start in terms of their understanding of the content.

It was found that between the pre-test and post-test there was no statistical significance between the treatment and control test scores, therefore the null hypothesis cannot be rejected. When looking at individuals that had a 60% increase in achievement, 84 percent of the treatment and 67 percent of the control showed that increase.

Students in the control group and the treatment group both showed improvement but the difference in improvement was not statistically significant, so the null hypothesis cannot be rejected. Alternatively, the treatment group did show more students with an improvement of at least 60% from the pretest to posttest but missed being a statistically significant difference by about one percent.

The internal consistency and reliability of the pretest and posttest was estimated using the Kuder-Richardson Formula 21 (KR21). The pretest reliability was 59% and the posttest reliability was 78%. These indices of test score consistency or precision translated into standard errors of measure of two points for each test. That is, a true theoretical test score is likely within two points plus or minus of an observed score. Internal consistency also assesses the degree to which a test measures one underlying trait. Reliability does not, however, identify the trait measured. Only validity can help determine what was measured by even a very precise test.

Table 1.
Pretests Control vs. Treatment

Method

μ_1: mean of Pretest when Sample = Control
μ_2: mean of Pretest when Sample = Treatment
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics: Pretest

Sample	N	Mean	StDev	SE Mean
Control	48	7.3750	3.3174	0.4788
Treatment	50	6.4600	3.1507	0.4456

Estimation for Difference

Difference	95% CI for Difference
0.9150	(-0.3835, 2.2135)

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
1.40	95	0.1651

Table 1 shows the comparison of the pre-test data between the control and treatment data using a *t*-test. The results show that there was not a statistical difference between the two groups as the p value was above 0.05.

Table 2.
Posttests Control vs. Treatment

Method

μ_1: mean of Posttest when Sample = Control
μ_2: mean of Posttest when Sample = Treatment
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics: Posttest

Sample	N	Mean	StDev	SE Mean
Control	47	14.0638	4.3758	0.6383
Treatment	49	13.7551	4.1909	0.5987

Estimation for Difference

Difference	95% CI for Difference
0.3087	(-1.4291, 2.0466)

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
0.35	93	0.7251

Table 2 shows that there was no statistical significance between the post-test scores of the control and treatment groups. A t-test was completed to determine that the p-value was 0.72 that is over the significant p value of 0.05.

Table 3.
Pre-to-Post Change Control vs. Treatment

Method

μ_1: mean of Pre_Post when Sample = Control
μ_2: mean of Pre_Post when Sample = Treatment
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics: Pre_Post

Sample	N	Mean	StDev	SE Mean
Control	47	6.7447	3.9643	0.5783
Treatment	49	7.2449	3.2630	0.4661

Estimation for Difference

Difference	95% CI for Difference
-0.5002	(-1.9760, 0.9756)

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
-0.67	89	0.5024

Table 3 demonstrates that there was no significance difference between the control and treatment samples in the change of mean between the pre and post-test. This means that there is not enough evidence to reject the null hypothesis based on the t-test p value of 0.50, which is higher than the p-value threshold of 0.05 for significance**.**

Figure 1.
Boxplot of Pretest Control vs. Treatment

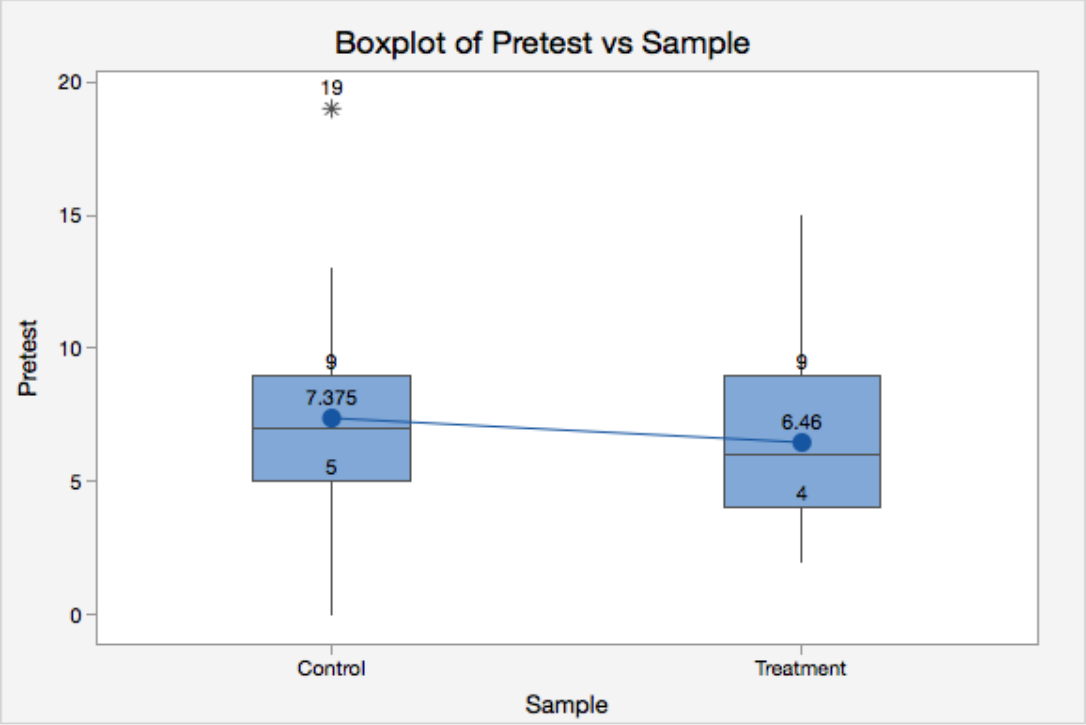

Figure 1 shows the distribution of the pre-test data for both the control and treatment groups and demonstrates that there is no significant difference and there is significant overlap between both groups.

Figure 2.
Boxplot of Posttest Control vs. Treatment

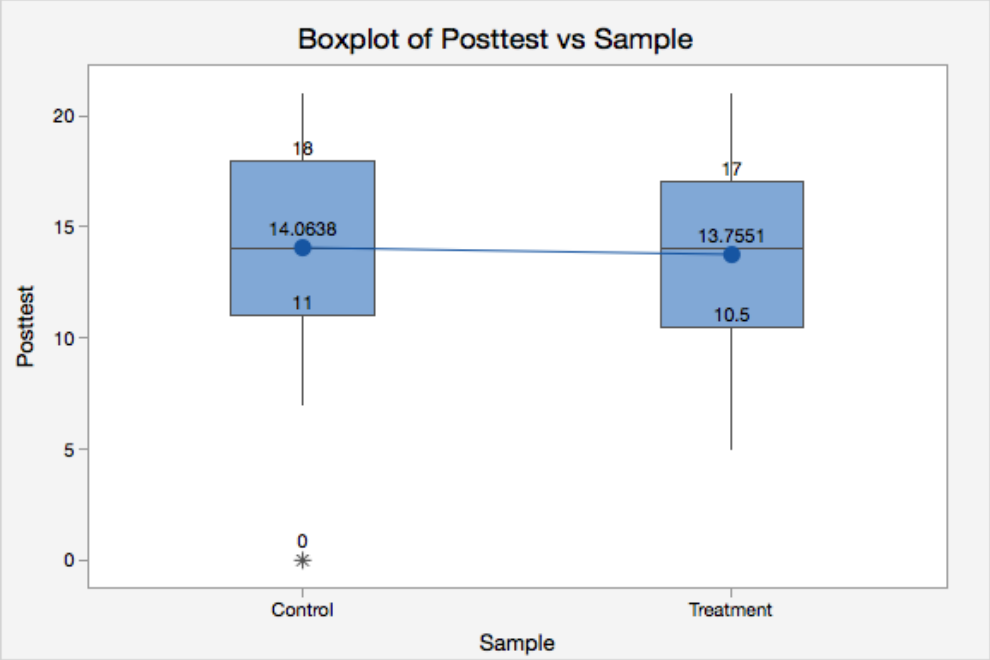

Figure 2 shows the distribution of the posttest data for both the control and treatment groups and demonstrates that there is no significant difference and there is significant overlap between both groups.

Figure 3.
Boxplot of Pre-to-Post Change Control vs. Treatment

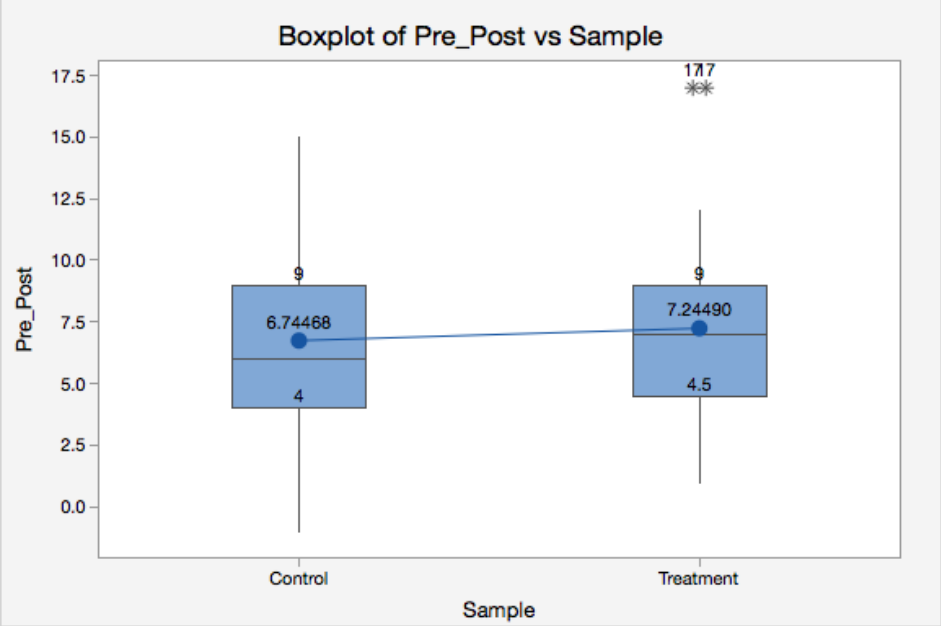

Figure 3 shows the distribution of the pre to mean change test data for both the control and treatment groups and demonstrates that there is no significant difference and there is significant overlap between both groups.

Table 4.
Gains from Pre-to-Post Score of More than 60%

Rows: Sample Columns: Recoded Sample From pct_gain

	<60%	>=60%	Missing	All
Control	15	31	13	46
	32.61	67.39		100.00
Treatment	8	41	9	49
	16.33	83.67		100.00
All	23	72		95
	24.21	75.79		100.00

Cell Contents: Count
% of Row

Chi-Square Test

	Chi-Square	DF	P-Value
Pearson	3.43	1	0.0641
Likelihood Ratio	3.46	1	0.0627

Table 4 shows a Chi-Square analysis to determine if an improvement of 60% from the pretest to the posttest was statistically significant between the control and treatment group. The results show a p-value of 0.06 based on the Chi-Square analysis. This did not meet the threshold of 0.05 so it is not statistically significant and there is not enough evidence to reject the null hypothesis.

In summary, the evidence from pre-and post-testing was insufficient to reject the null hypotheses of no mean population difference between the control and treatment groups after the treatment sample used the Project Tracker formative assessment strategy. The control and treatment samples were equivalent, on average, prior to the study and remained equivalent after the study was completed. The treatment sample did have more students improve by at least 60 percent from pre-to-post.

CHAPTER V

DISCUSSION

The original null hypothesis that the use of the Progress Tracker formative assessment would not lead to a statistically significant improvement of the mean scores compared to a control group could not be rejected. Additionally, the null hypothesis that the implementation of the Progress Tracker formative assessment will increase student achievement by 60 percent over a control group could not be rejected.

Implications of the Results

The results showed that the use of the Progress Tracker was not statistically better than not using the progress tracker when it came to keeping track of the mean for both groups. While it did not meet the threshold of 0.05, instead coming in at 0.06, the number of students that gained at least 60 percent from the pretest to the posttest was much closer to a significant value than looking solely at the mean.

While there is a lot of research about the effectiveness of formative assessments during instruction, this study was unable to show there was a significant difference between the treatment and control group. Changes will need to be made to the progress tracker and methods by which it is administered in order to determine if it is viable in some form or if there are more effective formative assessments available.

Theoretical Consequences

While there is sufficient evidence that formative assessments increase motivation and student achievement there was no statistically significant evidence that the Progress Tracker would differentially increase motivation or student achievement in this research. Much of the current evidence in support of such increases is focused on the core principles of formative

assessment including student involvement, feedback, learning progression and identification of the gap. Inclusion of these principles into specific strategies is not something that is discussed often considering the number of formative assessment strategies that can be found on websites and in professional development around the country.

The variety in formative assessment strategies and variability in implementation, formative assessment design and variability in student populations would need to be adjusted and repeated with other populations and the same populations to determine the effectiveness of this specific strategy.

Threats to Validity

Obtaining student buy-in with the strategy was difficult. Completion of the formative assessment was not graded and was treated by students as such. Some students did not complete the beginning and middle Progress Tracker or even the final Progress Tracker. The point value for the final Progress Tracker was small in point value. In the future, increasing engagement during all phases of the Progress Tracker might obtain better results. The Progress Tracker should also be used by other teachers and in other schools with different demographics.

While the feedback was instant after students completed the Progress Tracker, once they received the correct answers they were not tasked with taking the Progress Tracker again until a about a week later. Students did not have to continue after they received the feedback. Future iterations of the Progress Tracker should happen more often and task students with answering additional questions after getting a certain question wrong. Throughout the study, other in-class formative assessments were used to assess student understanding during the class and make adjustments to the lesson. This included polling students, whiteboards, exit tickets, sharing responses and discussion with peers and questioning techniques.

The Progress Tracker also included rating scales that were not graded which could have led to students not fully appreciating the multiple-choice questions that were graded and helped to keep track of student success. The multiple-choice questions were meant to emulate questions that students could see on teacher-designed assessments and state standardized assessments.

Connections to Previous Studies

Cauley and McMillan (2010) discuss how many specific formative assessment strategies are based on observational and anecdotal evidence to support the use of specific assessments. Not much is known about the specific benefits of certain commonly used formative assessments that are often seen in the classroom. This study set about determining quantitatively if the Progress Tracker formative assessment would lead to significantly improved achievement results for students but was unable to come to that conclusion.

Black and Wiliam (2010) discuss how formative assessments help to increase achievement but do not discuss specific strategies, instead describing formative assessments holistically. They note that formative assessments lead to a greater increase in lower achieving students. During this study it was found that the treatment group saw more students with a 60 percent increase over the control group but missed the significant threshold by one percent.

Hwang and Change (2011) and Sheard and Chambers (2014) used a technology-based approach that was similar to the approach of this study but their technology-based formative assessment was complex enough to offer students different questions based on their response to previous questions instead of offering the correct responses right away as they were in this study. This is one possibility, as well as the other from the previous section, that could have led to no difference between the control and treatment groups.

Conclusion

Specific formative assessment strategies are often used widely in the classroom, but more research needs to be done to look into their effectiveness, design, and implementation. Most evidence is anecdotal or observational surrounding specific strategies. This study set out to look at the quantitative effectiveness of the Progress Tracker formative assessment on student achievement. It was found that there was no significant difference between the control and treatment group. This study did find that more students had an increase of at least 60 percent in their assessment scores in the treatment group but missed the significance threshold by one percent.

It was critical to have a control group that was similar to the treatment group at the start of the study because without the control group the treatment would have showed a significant increase between the pretest and posttest for the treatment group. However, when comparing the increases in mean between the control and treatment group, it was evident that there was no significant difference between the control group that did not receive the Progress Tracker and the treatment group that did. If there were no comparison group in this study, it would have yielded very different conclusions.

This rationale presented above suggests that it was another factor that contributed to the increase in means for both groups. It could have been the other formative assessments used during instruction in the classroom for all classes, direct instruction of the content, collaborative learning or the learning activities completed during the class that contributed to the outcome. Many teaching strategies and techniques are used during instruction that could have led to the increase in mean score for both groups. With further studies, the strategies used during the unit of content could be tested in a similar way that the Progress Tracker was tested to determine

which teaching strategy or other formative assessments caused the significant increase in mean score for both these groups.

Formative assessments remain an evidence-based strategy to improve motivation, engagement and achievement but the Progress Tracker formative assessment would need adjustments in design and implementation as well as repeated testing before it could be an effective assessment in the classroom.

REFERENCES

Andersson, C., & Palm, T. (2017). The impact of formative assessment on student achievement: A study of the effects of changes to classroom practice after a comprehensive professional development programme. *Learning & Instruction, 49.*

Ash, D., & Levitt, K. (2003). Working within the zone of proximal development: Formative assessment as professional development. *Journal of Science Teacher Education, 14*(1), 23-48.

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment: Formative assessment is an essential component of classroom work and can raise student achievement. *Phi Delta Kappan, 92*(1), 81.

Carrillo-de-la-Peäna, M., Bailláes, E., Caseras, X., Martâinez, A., Ortet, G., & Pâerez, J. (2009). Formative assessment and academic achievement in pre-graduate students of health sciences. *Advances in Health Sciences Education, 14*(1), 61-67.

Cauley, K., & McMillan, J. (2010). Formative assessment techniques to support student motivation and achievement. *The Clearing House, 83*(1), 1-6.

Clark, I. (2011). Formative assessment: Policy, perspectives and practice. *Florida Journal of Educational Administration & Policy, 4*(2), 158-180.

Crichton, H., & McDaid, A. (2016). Learning intentions and success criteria: Learners' and teachers' views. *Curriculum Journal, 27*(2), 190-203.

Crumrine, T., & Demers, C. (2007). Formative assessment redirecting the plan: High school science teachers study and implement formative assessment strategies. *The Science Teacher, 74*(6), 64-68.

Faber, J., Luyten, H., & Visscher, A. (2017). The effects of a digital formative assessment tool on mathematics achievement and student motivation: Results of a randomized experiment. *Computers & Education, 106*.

Heritage, M. (2007). Formative assessment: What do teachers need to know and do? Phi Delta Kappan, 89(2), 140-145.

Hudesman, J., Crosby, S., Flugman, B., Issac, S., Everson, H., & Clay, D. (2013). Using formative assessment and metacognition to improve student achievement. *Journal of Developmental Education, 37*(1), 2-4.

Hwang, G., & Chang, H. (2011). A formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students. *Computers & Education, 56*(4), 1023-1031. doi:10.1016/j.compedu.2010.12.002

Martínez-Gudapakkam, A., Mutch-Jones, K., & Hicks, J. (2017). Methods and strategies: Formative assessment practices to support students who struggle in science. *Science and Children, 55*(2), 88-93.

Nolen, S. (2011). The role of educational systems in the link between formative assessment and motivation. *Theory into Practice, 50*(4), 319-326.

Ozan, C., & Kincal, R. (2018). The effects of formative assessment on academic achievement, attitudes toward the lesson, and self-regulation skills. *Educational Sciences: Theory and Practice, 18*(1),

Sheard, M., & Chambers, B. (2014). A case of technology-enhanced formative assessment and achievement in primary grammar: How is quality assurance of formative assessment assured? *Studies in Educational Evaluation, 43*.

Tomanek, D., Talanquer, V., & Novodvorsky, I. (2008). What do science teachers consider when selecting formative assessment tasks? *Journal of Research in Science Teaching*, 45(10), 1113-1130. doi:10.1002/tea.20247