2017 HAWAII UNIVERSITY INTERNATIONAL CONFERENCES
ARTS, HUMANITIES, SOCIAL SCIENCES & EDUCATION JANUARY 3 - 6, 2017
ALA MOANA HOTEL, HONOLULU, HAWAII

REFLECTIVE ASSESSMENT, FEEDBACK AND HIGH SCHOOL ACHIEVEMENT

BALIRAM, NALLINE
SEATTLE PACIFIC UNIVERSITY
SCHOOL OF EDUCATION

Dr. Nalline Baliram
Seattle Pacific University
School of Education

Reflective Assessment, Feedback and High School Achievement

Synopsis:

This study investigated the effects of reflective assessment linked with content specific feedback on student learning. The findings in the study support previous studies that these strategies identified as best practices can positively impact student learning.

Reflective Assessment and Feedback in High School Mathematics

Nalline Baliram
Assistant Professor
School of Education
Seattle Pacific University, Seattle, WA

Abstract

The purpose of this study was to determine the effects of metacognitive strategies and content-specific feedback on student achievement in high school mathematics. Participants in the study consisted of a convenience sample of 75 honors geometry students in grades 9 and 10 in a private high school located in Florida. A quasi-experimental, nonequivalent control-group design was employed in the study. Parametric methods were used to address the research question. Specifically, an ANCOVA two-tailed test was performed. There was a statistically significant difference between groups in the post-test. More importantly, students in the experimental group outperformed the students in the comparison group.

The findings of this study offer a modest contribution to the body of empirical research on the impact of metacognitive practice and content-specific feedback on academic achievement at the high school level. Further studies are warranted to add to the body of literature and more specifically to provide great clarity regarding the magnitude of the current investigation.

Introduction

The education environment is in an era of high-stakes testing and heightened pressure to improve student achievement. Teachers are increasingly expected to help their students produce favorable outcomes on high-stakes standardized tests (Guth et al., 1999). Therefore, teachers continue to explore methods, concepts, and strategies that have grounds to help their students acquire, make sense of, and retain knowledge. Reflective assessment, a longtime strategy, can be utilized as a diagnostic approach to provide feedback to both the teachers and students over the course of instruction. When students are able to practice reflection, teachers are able to identify areas where students are struggling and further provide feedback in an attempt to ameliorate the situation.

An extension of reflective assessment is feedback, one of the most powerful influences on learning and achievement (Hattie & Timperley, 2007). Feedback can be perceived to be positive or negative, therefore, the type of feedback and the way it is given can be differentially effective. Both reflective thinking and teacher feedback can be characterized as highly esteemed and widely used techniques that are utilized in a variety of professions to aide in adapting and making decisions.

Research suggests formative assessment including reflective practice and feedback are tightly linked with instructional practices. Therefore, teachers must consider how their classroom activities, assignments and tests support student learning and allow students to freely communicate what they know, what they can do and areas in which they continue to struggle. Teachers must then use this information to improve teaching and learning.

Definitions

Metacognition. The term metacognition appeared as an interesting and promising new area of study based on psychologist John Flavell's work several decades ago (Flavell, 1979), and is often associated with reflective assessment. In education, metacognition refers to the way teachers and students plan, monitor and assess understanding and performance. Metacognition involves thinking about one's cognitive activities, and this skill allows the individual to organize, monitor, evaluate, and regulate the thinking process (Bandura, 1997). It includes thinkers being aware of how they think and learners being aware of how they learn. Ultimately, as literature suggests, metacognition is defined as "thinking about thinking" (Costa, 2001; Flavell, 1979; Schoenfeld, 1987). In this sense, student reflection represents a value-added component often missing in teaching and learning.

Reflective assessment. John Dewey (1910) wrote that reflection has the potential to happen when there is a feeling of doubt or perplexity. He defined reflection as involving a consecutive order so that each idea determines the next outcome. Defined by Leung and Kember (2003), *reflection* is an attempt to understand an issue or question within a personal context or going beyond learning to assimilate information to make meaning (p. 64). *Assessment* is a valid measure of learning that provides feedback to both the teacher and the learner for the purpose of improving teaching and learning (Popham, 2014;

Stiggins, 1996). Therefore, *reflective assessment* implies active contemplation on the cognitive process of knowledge, skills, situations or experiences with some kind of measurement, typically formative. In this sense, reflective assessment by students and teachers is assessment for purposes of learning and growth.

Feedback. Formative assessment refers to assessment that is specifically intended to generate feedback on a student's performance with the intent to improve learning (Black & Wiliam, 1998; Sadler, 1998). John Hattie (2012) theorized that the most powerful strategy that enhances achievement is feedback. Irons (2008) defined feedback as "any information, process or activity which affords or accelerates student learning based on comments relating to either formative or summative assessment activities" (p. 7). According to Brookhart (2008), effective feedback should be clear, age-appropriate, content specific, timely, and of high quality. Typically, feedback comes from teacher to student; however, effective feedback can also come from student to student as well as student to teacher.

Theoretical Framework

Metacognition and Reflective Assessment. Reflective thinking became a vital theme during the progressive movement in American education. Dewey (1910), considered reflection an "active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it, and the further conclusions to which it tends" (p. 6). He further theorized reflection as a process that enables the learner to move from one experience to the next and which involves a deeper understanding of its relationships with prior experiences and ideas. It is a rigorous and systematic way of thinking with its roots in scientific inquiry.

Flavell (1979), who defined metacognition as 'thinking about thinking', acknowledged the significance of metacognition in a wide range of applications which included reading, oral skills, writing, language acquisition, memory, attention, social interactions, self-instruction, personality development and education. Components of metacognition can be activated intentionally; this could be through a memory search with the purpose of retrieving specific information (Flavell, 1979). Such components can help the individual make meaning and discover behavioral implications of metacognitive experiences.

As noted by Jerome Bruner (1961), students should spend more time studying problems in depth and less time covering a wide range of topics; meaning students be allowed opportunities to consistently practice formative self-assessment as a means of clarifying their thinking about what they are learning. Bandura (1997), an advocate of the significance of self-efficacy in learning, proposed that effective intellectual functioning requires metacognitive skill to organize, monitor, evaluate, and regulate the thinking process. He wrote that "Metacognition involves thoughts about one's cognitive activities rather than simply higher order cognitive skills" (p. 223).

Based on these theoretical perspectives, it is clear that the idea of teaching students to think about their own thinking has been in existence for a long time, eventually taking its

place as a vital theme of the progressive movement. Reflective thinking relates to constructivism in that learners are conscious of how they learn and can therefore regulate their progress (Joyce, Weil, & Calhoun, 2009). This suggests when students are taught science for example, teaching the scientific thinking processes is not enough. A value-added component occurs when students can demonstrate to themselves their academic achievement, when they are able to reflect on their own learning and the learning process (Marzano et al., 2012). The theoretical works of Dewey (1910), Bruner (1961), and Bandura (1997) continue to inform education today specifically in the area of metacognitive practice.

Feedback. In the 1960s, psychologists argued that schools could improve instruction by adopting a more systematic approach and therefore, borrowed the idea of feedback from engineering systems theory. However, through intervention, they discovered feedback must be designed to be a part of a system instead of just telling students whether their responses were correct or incorrect (Wiliam, 2012). Therefore, when feedback is given, it should inform the students that the current performance falls short of the learning goal or the goal has already been reached.

Hattie and Timperley (2007) in their article, *The Power of Feedback*, hypothesized feedback as one of the most powerful influences on learning and achievement. The authors described feedback as a "consequence" of performance (p. 81) suggesting successful outcomes can occur when student are able to make meaning of the feedback they receive. According to Irons (2008), feedback can benefit students only if it indicates clear expectations that can be attainable, compares their current level of performance with the intended level, and provides specific actions students must follow in order to improve their learning or to close the gap between their current to intended levels of performance.

The theoretical background demonstrates that reflective assessment or metacognitive practice as well as feedback are significant components of formative assessment that have serious implications for teaching and learning. Students should not just learn; they must be encouraged to reflect on how they learn and implications for generating success (Marzano et al., 2012). Furthermore, teachers must be able to recognize valuable insights in their students' metacognitive practice and provide them with meaningful feedback that will enhance their learning (Hattie & Timperley, 2007). Establishing a teacher-student relationship, rooted in these theoretical underpinnings, can foster a classroom climate in which feedback and reflective assessment are prevalent and highly valued.

Both types of formative assessments, reflective assessment and feedback, offer promise of success when they are focused on students' performance and their ability to accomplish their academic goals. A number of studies have investigated strategies that allow students to reflect on their thinking and learning as well as stimulate the metacognitive process to occur. Additionally, a number of studies have examined the impact of teacher feedback on student learning at a variety of ability levels (Butler & Nisan, 1986; Nunez et al., 2013; Siewert, 2011).

Purpose and Rationale

The body of knowledge regarding metacognition or reflective assessment is rich. However, there are few studies that link student reflection with teacher feedback and examine the effects of both practices on academic achievement. This study seeks to shift the priority of instructional delivery by validating the need to link reflective assessment with feedback. The more evidence that teachers receive about the benefits of these two approaches, more likely that they will enhance their learning environment by integrating the approaches in their daily practice. Research conducted in a realistic classroom environment will provide teachers with helpful information that will inform their instructional practice.

The purpose of this study was to bridge the research gap regarding the use of metacognitive practice and feedback as part of the daily teaching and learning routine for teachers and students. This study examines the use of metacognitive practice (reflective assessment) and teacher feedback during geometry instruction at the high school level. The following research question guided the investigation is: (1) What are the effects of practicing reflective assessment and providing teacher feedback on achievement of high school geometry students? The hypothesis tested is as follows: There is a statistically significant difference for Group (two levels: reflective/feedback and non-reflective/feedback) on academic achievement of high school geometry students as measured by their score in the end of unit assessment.

Research Methodology

This study examined the effects of metacognitive strategies and content-specific feedback on the academic achievement of high school students in mathematics. Specifically, the researcher examined the effects of situated metacognition, in the form of reflective assessment, linked with feedback on high school students studying Honors Geometry. According to the Florida Department of Education, Geometry is a required course for all high school students. Honors Geometry is an advanced class designated for high achieving math students who want to learn at an accelerated pace and deeper level. The researcher adapted the specific intervention and procedure focused on integrating reflective assessment from several prior studies (Bianchi, 2007; Bond, 2003; Evans, 2009).

Research Design

An analysis of covariance (ANCOVA), nonequivalent control-group design, two-tailed test was employed in the study. The independent variable examined was the use of reflective assessment linked to feedback. The reflective prompts utilized in this study were based on those articulated by Ellis (2001) and Mevarech and Kramarski (1997). The dependent variable was the performance on a measure of geometry content covered over the course of the intervention. The criterion instrument was developed by the publisher of the geometry textbook utilized in the course and aligned with the geometry content knowledge (Hall, Kennedy, Bass, & Wiggins, 2012). The specific content of the

instrument related to the *Pythagorean Theorem and Special Triangles* unit, a required topic taught in all geometry courses in Florida. The criterion instrument was administered prior to the intervention (pretest) and at the completion of the intervention (posttest).

The procedure lacked random selection of participants to treatments due to the constraint of preexisting classroom assignment for students. Despite this, the intact student groups were randomly assigned to either the comparison group or the experimental group. In total, this involved random assignment of five intact classes, with each class containing approximately 17 students. The characteristics of the participating student population are discussed later in this chapter.

A two-tailed test was selected for the following reasons: (1) the researcher wished to achieve a more rigorous test result, (2) a two-tailed test will provide the researcher with a "safeguard" against unexpected results, and (3) the researcher opted for a non-directional hypothesis in response to the research question (Cho & Abe, 2013). A pre-test was administered because random selection at the level of participants was not feasible. The pre-test scores were utilized to elicit potential pre-existing differences between students in the experimental and comparison groups and therefore named as the covariance in the data analysis.

According to Field (2013), an ANCOVA is used when pre-existing differences need to be controlled. However, with this design, the testing poses threats to both internal validity and external validity. Regarding internal validity, the same instrument was administered to the participants on two different occasions. Because students were familiar with the assessment, there was potential for gains in the students' scores across tests (Campbell & Stanley, 1963). This phenomenon is described as students becoming "test-wise" (Gall, Gall, & Borg, 2003). To address this threat to internal validity, both the experimental and comparison groups received equivalent exposure to the instrument, which should therefore minimize the differential effects of testing between the two groups.

Regarding external validity, it is possible that the assessment would interact with the intervention in such a way that it could enhance the effect of the treatment, which is known as test sensitization (Gall et al., 2003). The researcher acknowledged the possibility that the administration of a pretest or post-test could activate the students' awareness of their attitudes toward the concept which could sensitize them to react to the content and intervention in a way that would affect the outcome.

Participants

A convenience sample was used to recruit participants at the classroom level. The sample consisted of students from five intact Honors Geometry classes taught by the same teacher in a private high school located in an urban city in Valousa County, Florida. According to the 2014 census, the city in which the school is located reported the following demographic data: 48.9% male and 51.1% female; ethnographic data include 8.7% Hispanic, 0.2% American Indian, 2.6% Asian, 31.7% African American, 54.2%

Caucasian, and 2.6% Multi-ethnic. Furthermore, approximately 16.8% of the town's population is below the poverty line.

The school is a private, Catholic high school that serves students who come from both private and public middle schools in the area. At the time of the study, the school population was represented by over 50 different zip code areas within the county and consisted of 474 students in grades 9 to 12. The school consisted of 48% male and 52 % female students. The ethnographic makeup of the students is as follows: 5% Hispanic, 10% Asian, 7% African American, 75% Caucasian, and 3% Multi-ethnic. Furthermore, 70% of the students received tuition assistance to help families bridge the gap between what they can afford to pay and the tuition cost. The breakdown of this assistance was as follows: 14% of the student are financially supported with one fourth of tuition assistance, 21% with one fourth to one half of tuition assistance, 7% with one half to three-fourths of tuition assistance and 28% with more than three-fourths of tuition assistance. Finally, 6% of the student body received some form of remedial accommodation based on their learning disabilities.

At the classroom level, the participating teacher reported the demographic data for students participating in this study: from a sample size of 75, 45.3% male and 54.7% female; the ethnographic data included 8.0% Hispanic, 10.7% Asian, 1.3% African American, 61.3% Caucasian, and 10.7% Multi-ethnic and 8% other. Additionally, 85.3% represent the grade 9 class and 14.7% are grade 10 students. These data are presented in Table 1.

Table 1
Demographic Information of Sample

		Frequency	Percentage
Ethnicity			
	Hispanic	6	8.0
	Asian	8	10.7
	African American	1	1.3
	Caucasian	46	61.3
	Multi-ethnic	8	10.7
	Other	6	8.0
Gender			
	Male	34	45.3
	Female	41	54.7
Grade Level			
	9^{th} Grade	64	85.3
	10^{th} Grade	11	14.7

The overall sample size for Honors Geometry students was 75 consisting of five classes of approximately 17 students in each class. These five intact classrooms were randomly assigned to one of the two groups by a "draw from the hat" process, which was observed by two individuals unaffiliated with the study. Of the five classes, two were randomly assigned to the experimental group and three to the comparison group. The total group

sizes were 33 students in the experimental group and 42 students in the comparison group. Table 2 provides characteristics of the sample by gender and group.

Table 2

Sample by Gender

Grouping Variable	Male	Female	Total
Comparison Group	18	24	42
Experimental Group	16	17	33
Total	34	41	75

This study occurred over four weeks consisting of 16 instructional days. All five classes received instruction for ten class periods each 45 minutes long. The remaining six instructional days were identified as block periods in which the class periods were 90 minutes long and students attended half of the total number of classes each day. During the course of the study, each student attended ten 45 minute classes and three 90 minute classes.

Instrumentation

The criterion instrument used for data collection in the pretest and post-test was the unit test developed by publishers of the Geometry textbook utilized as a resource (Hall et al., 2012). The specific content of the instrument related to the *Pythagorean Theorem and Special Triangles* unit, the Geometry topic that students focused on during the study.

To examine the reliability of the instrument, the researcher conducted a test-retest analysis using the post-test and retention test. According to recommendations by Gall et al. (2003), a correlation coefficient is calculated to determine the reliability of the test scores, a direct measure of consistency, on the same measure between two different occasions. The bivariate coefficient between the pretest and the posttest revealed a coefficient of stability of $r = .47$, which is statistically significant at the $p < .01$ level. The measure of internal consistency was calculated using Cronbach's alpha. The Cronbach's alpha is .75, which indicates a high level of internal consistency (Field, 2013).

Procedure

The intervention lasted four weeks, which aligned with the district-specified timeline for the *Pythagorean Theorem and Special Triangles* unit. Six of the instructional sessions were 90 minutes in length. The remaining 10 sessions were 45 minutes in length. Toward the end of the study, the participating teacher was absent for one day. Although the teacher provided an assignment that was aligned with the topic, the students were not exposed to the traditional instruction. This anomaly could present itself as a potential threat to internal validity so, to address this threat, all five classes were without a math instructor for a total of one day.

One comparison class and one experimental class met in the early morning while the remaining three classes met after lunch. This difference in the time of day potentially represents a confounding variable that constitutes a threat to internal validity (Gall et al., 2003). The time of day classes were held and the random assignment of the classes to the groups was beyond the control of the researcher.

Intervention. On the first day of the study, students in both groups were administered a pretest. During the remaining time, all of the students received equivalent instruction in the *Pythagorean Theorem and Special Triangles*, with one exception: The experimental group completed the metacognitive prompts along with two practice problems and received content-specific feedback for a total of 12 instructional periods during the last five to 10 minutes of the instructional period.

The students in the experimental group recorded their responses on a note card, which was then collected by the teacher. This process constituted reflective assessment. During this time, the comparison group either reviewed the learning target for that day or began their homework assignment. The following metacognitive prompts provided to students in the experimental group included:

1. I Learned Statement (Ellis, 2001): *Today, I learned …*
2. Strategic Questioning (Mevarech & Kramarski, 1997): *I can now apply ... to solve …*
3. Clear and Unclear Windows (Ellis, 2001): *I understand ... but still don't understand …*

The participating teacher de-identified the reflective cards and made them accessible to another Geometry teacher (not affiliated with the participants) to provide the feedback. This procedure was done to avoid bias and to allow students to receive content-specific feedback on the reflective assessment card in response to any specific questions or comments each student posed. Additionally, common trends such as misconceptions in areas where most students showed they struggled were identified and communicated to the participating teacher.

When the reflective cards with specific content feedback were returned to the experimental group at the beginning of the class, the teacher provided additional general feedback that would improve students' understanding of the concept. In contrast, the comparison group began class by practicing problems to review their prior knowledge. Quality feedback, in terms of content-specific and general can influence instructional revision in a positive sense when it is immediate and focused on student reflection and learning (Black & Wiliam, 1998; Guskey & Marzano, 2003; Hattie, 2012).

Data Analysis

The researcher used SPSS version 23 general linear model to address the hypothesis. Prior to computing inferential statistics, the data were scanned for missing scores as well as any outliers. One case was identified in which the student was missing a pretest score. To address this, the missing case was replaced with the mean score of the pretest. Data

were analyzed to ensure parametric procedures would be appropriate. Table 3 provides the descriptive for the pretest and post-test. Each of these variables represents a separate administration of the same instrument, *Pythagorean Theorem and Special Triangles* Test. The possible range of scores on the instrument was 0 to 100. Tables 4 and 5 present the data disaggregated by group assignment for each variable.

Table 3

Descriptive Statistics for Pre-Test and Post-Test

	N	Mean	*SD*	Skewness	Kurtosis
Pre-Test	75	19.15	10.55	.68	.92
Post-Test	75	85.27	13.41	-1.29	1.31

Table 4

Descriptive Statistics for Pre-Test

Group	*N*	Mean	*SD*	Skewness	Kurtosis
Comparison	42	17.31	9.95	.15	-.78
Experimental	33	21.48	10.98	1.17	1.82

Table 5

Descriptive Statistics for Post-Test

Group	*N*	Mean	*SD*	Skewness	Kurtosis
Comparison	42	81.90	14.84	-1.14	.64
Experimental	33	89.55	10.01	-1.27	1.29

The pretest was administered prior to the study. The data presented in Table 4 suggest that students knew very little of the unit content prior to the intervention. Additionally, both skewness and kurtosis statistics for the pre-test distribution fall within plus or minus one. This suggests the data for the pre-test produced a normal distribution.

The post-test was administered at the completion of the intervention. According to the data presented in Table 5, post-test scores suggest a ceiling effect occurred. The mode reported for the post-test was 99, which is one point below the maximum possible score. Both the skewness and kurtosis statistics for the post-test distribution fall outside the range of plus or minus 1, which suggest a non-normal distribution (Gall, et al., 2003). This was confirmed by both the Kolmogorov-Smirnov and Shapiro-Wilk tests' of normality. The skewness statistics of -1.29 (*SE* = .27) indicates a negative skew to the data and the kurtosis statistics of 1.31 (*SE* = .54) shows a peak in the data. The ceiling effect is a possible explanation for the negative skew and the mode in the post-test explains the kurtosis statistics. Figure 1 provides an illustration of the distribution of the post-test data.

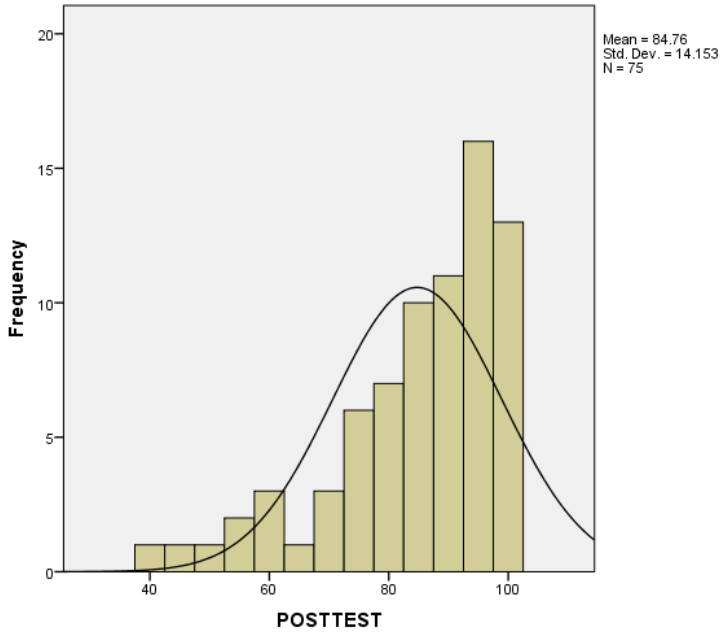

Figure 1. Post-Test Scores

Results

Based on the research question, the researcher wanted to determine if there is a statistically significant difference in academic achievement of high school geometry students who practice reflective assessment and receive content specific feedback and those who do not practice reflective assessment. The research question was tested using an analysis of covariance (ANCOVA). The significance of effects was analyzed at an alpha level of .05 (Gall et al., 2003).

The ANCOVA two tailed test indicated that there was significance in the main effect of the variable group ($F(1, 73) = 4.11$, $p < .05$, $\eta_p^2 = .05$). This suggests that at an alpha level of $p < .05$, there was a statistically significant difference between groups (comparison and experimental) on the end of unit assessment. The experimental group significantly outperformed the comparison group in the post-test ($p < .05$, Cohen's $d = .60$).

Discussions and Limitations

It was the intent of the researcher to apply the theories of metacognition and content-specific feedback that have been articulated by Bandura (1997), Dewey (1910), Flavell (1977), Hattie and Timperley (2007) and Vygotsky (1978) to the classroom setting. Beyond answering the specific research question, an additional aim in this study was to contribute to the growing body of knowledge pertaining to effective ways to use metacognitive instruction and provide effective content-specific feedback to improve student achievement and learning.

The results of this study offer tentative support for reflective strategies linked with content-specific feedback embedded as formative assessments in daily activities. Because of a lack of studies explicitly linking both strategies, it is premature to confirm any effects the intervention had in the learning environment. Previous studies focused either on reflective assessment or teacher feedback have shown that these approaches have positive effects on student achievement (Bond & Ellis, 2013; Butler & Nisan, 1986; Evans, 2009; Kramarski & Mevarech, 2003; Nunez et al., 2015). However, because the cited studies did not explicitly link reflective assessment with content-specific feedback, it is difficult to conclude the findings support the work of previous studies. For this reason, further research is warranted to examine the effectiveness of metacognitive practice linked with content-specific feedback on academic achievement.

In terms of setting and reflective assessment, the present study is similar to the work of Evans (2009) whose sample consisted of grade 9 high school English language students. In both Evans' (2009) and the present study, the researchers focused on the effectiveness of reflective assessment when used daily. In the present study, the researcher randomly assigned five intact classes to one of two groups, comparison and experimental, while Evans (2009) randomly assigned nine intact classes to one of three groups, control, comparison and experimental. The data gathered from both studies revealed a statistically significant difference in the students' achievement scores favoring the experimental group on both the post-test and retention test.

Three significant differences exist between the present study and the study conducted by Evans (2009). First, in the present study, the researcher used a nonequivalent control-group design with repeated measures. A pre-test was administered to the students to compensate for the non-equivalent group design and though the experimental group outperformed the comparison group, there was no statistically significant difference between both groups in terms of ability. The data from the pre-test showed that students in both groups knew very little of the unit content prior to the intervention. Evans (2009) used a post-test only control group design, which indicated that a pre-test was not administered to the students.

Second, in the present study, both the experimental and comparison groups consisting of five intact classes were instructed by the same teacher. In the study conducted by Evans (2009), three participating teachers each taught three of the nine intact classes. Although the classes were randomly assigned to the teachers, the three instructional styles could have affected the outcome of the study. This potential threat to internal validity is called selection bias (Campbell & Stanley, 1963).

Third, in the present study, the participating teacher collected the reflective cards at the end of each period and after de-identifying them, gave them to another mathematics teacher to provide the students with content-specific feedback, which were then returned to the students for further review. In the study conducted by Evans (2009), the participating teachers collected and maintained the written responses to the prompts from the students. The researcher then collected the student work during class visits and maintained them as part of the record-keeping for the study. It must be noted that in the

study conducted by Evans (2009), no explicit feedback based on the written responses to the reflective prompts was provided to the students.

In terms of feedback, the present study is similar to the work of Nunez et al. (2015) who examined the relationship between teachers' feedback on homework and academic achievement. The present study was conducted in a high school class in the United States and focused on geometry students primarily from grade 9. The study conducted by Nunez et al. (2015) focused on students from grades 5 – 12 in three schools in northern Spain, and the researchers found teachers' feedback on homework was positively and significantly related to the quality and amount of homework the students completed. Additionally, the quality and amount of homework completed positively and significantly predicted academic achievement.

Two significant differences exist between the present study and the study conducted by Nunez et al. (2015). First, in the present study, written feedback was content-specific and provided based on the students' daily reflective assessment as opposed to Nunez et al. (2015) who provided feedback on students' homework, which was not identified as a form of reflective writing. Additionally, the type of feedback provided in the Nunez et al. (2015) study was a letter grade based on completion along with positive reinforcement. It is unclear if the positive reinforcement was written or oral. Though this quality of feedback positively and significantly predicted academic achievement, it was not content-specific.

Second, in the present study, the researcher utilized a quasi-experimental design and analyzed the data using an ANCOVA, two-tailed test. The purpose of this type of test was to determine the impact the intervention had with the experimental group when compared with the comparison group over time. In the study conducted by Nunez et al. (2015), a structural equation model (SEM) was analyzed using AMOS 18 (Arbuckle, 2009) to test the relationship between the teachers' feedback as perceived by the students, the homework-related variables, and student achievement.

A limitation with regard to the participants, was that a convenience sample was employed in the present study. A related matter is that the use of a convenience sample raises a threat to external validity, specifically in terms of population validity. When a sample such as the one made available for the present study is not necessarily reflective of a broad population, inferential statistics should be used with caution when certain conditions are not met (Gall et al., 2003). To address this issue, several characteristics of the sample were provided in Chapter Three including details pertaining to the participants in the study, the sample they were drawn from, and the defined population.

Another limitation, known as the Hawthorne effect, raises a threat to external validity. The Hawthorne effect occurs when individuals are aware that they are participating in an experiment (Gall et al., 2003). The nature of the current study required assent from the students, thus raising the possibility of the Hawthorne effect. Therefore, the external validity of the treatment was potentially compromised and encumbers the ability to generalize the findings.

Three potential limitations with respect to the study's methodology surfaced. First, the same criterion instrument was used for the pre-test, post-test and retention test. With regard to internal validity, this is a possible concern associated with testing effect (Campbell & Stanley, 1963). Because students were familiar with the assessment, there was potential for gains in the students' scores across tests, which is known as becoming "test-wise" (Campbell & Stanley, 1963; Gall et al., 2003). To alleviate this threat to internal validity, both groups received equivalent exposure to the instrument, thus minimizing the differential effects. Regarding external validity, it is possible that pre-test and post-test sensitization occurred. Sensitization occurs when the pre-test serves as a learning experience on its own, which has meaningful impact on the treatment. This potential interaction of testing with the treatment hinders the ability to generalize from the study's findings (Gall et al., 2003).

Another potential limitation was the timeliness of the feedback provided to the students. Feedback was provided within three to five days from the time the students completed their reflection cards. Since researchers suggest feedback should be provided in a timely manner (Hattie, 2012, Hattie & Timperley, 2007), the delay of feedback could have negatively impacted the validity of the test scores. However, according to Brookhart (2008), slightly delayed feedback can be meaningful as long as it is provided while students are mindful of the learning goal and content.

Third, the researcher provided the feedback to the students. Although the purpose of this protocol was to strengthen the validity of the study, it raises the concern of experimenter bias. Experimenter bias occurs when the researcher unintentionally influences the results to produce a certain outcome (Gall et al., 2003). However, because the researcher was a former geometry teacher and not affiliated with the participants, the feedback provided was content-specific and therefore, avoided other potential biases related to the personal knowledge of the students.

Implications of the Findings and Suggestions for Future Research

The findings from this study offer a modest contribution to the body of empirical research on the impact of metacognitive practice and content-specific feedback on academic achievement at the high school level. Further studies are warranted to add to the body of literature and more specifically to provide greater clarity regarding the magnitude of the current investigation. Although the findings from the study show moderate effect sizes, based on the limitations referenced in this article and lack of studies that link student reflection with teacher feedback, the researcher recommends further study to support any broad-based conclusions. To date, the majority of studies have focused on either the impact of reflective assessment or the impact of teacher feedback (Bond, 2013; Butler & Nisan, 1986; Evans, 2009; Kramarski & Mevarech, 2003; Lew & Schmidt, 2011; Nunez et al., 2015; Siewert, 2011; Zan, 2000).

Future studies should be crafted to include larger carefully selected samples across diverse settings to examine the effects of reflective assessment linked with content-

specific feedback on academic achievement and to probe its validity and usefulness for a broader population. This could include conducting studies across a variety of disciplines and grade levels from elementary to college level with the intent to more clearly develop a portrait of how the use of reflective assessment, when linked with teacher feedback, impacts learning and retention. Additionally, studies that employ the use of various designs and analyses are necessary to yield more generalizability. Although studies involving a convenience sample can provide valuable insights, "repeated replication of the findings is much stronger evidence of their validity and generalizability than is a statistically significant result in one study" (Gall et al., 2003).

Apart from conducting studies with the intent to examine broad-based effects of metacognitive practice and feedback, studies that examine discrete aspects of implementation are also recommended. For instance, advocates of reflective practice argue it is a skill that must be taught in order to be utilized effectively (Borich, 2014; Costa, 2001; Ellis, 2001). Perhaps, similar to the study conducted by Kramarski and Mevarech (2003), future studies should involve in-service teacher training that focuses on pedagogical practices involving metacognition. Furthermore, explicit metacognitive practice should be addressed in the classroom.

Additionally, it would benefit teachers to understand what qualifies as good feedback and decide how it should be given based on students' abilities, learning needs, and interests (Brookhart, 2008). Another topic of interest for future studies is the timeliness of feedback. When effective feedback is timely, it enables the students to process and implement the feedback (Brookhart, 2008). In turn, students become more receptive to the feedback while they are still mindful of the topic, assignment, or performance in question. In other words, feedback should be given when there is still time to correct errors. Otherwise, when it is no longer relevant to the current or future content, the feedback is pointless (Kulik & Kulik, 1988).

Implications for Classroom Practice

The effect sizes reported show that the results obtained carry practical significance for both teachers and students in classroom environments. This strategy that involved class closure in the form of reflective assessment may have positively affected what the students learned and the depth at which they learned it, when content-specific feedback was provided to each student.

Based on the findings of the current study, the researcher concludes that formative assessment, when linked with content-specific feedback, led to improved learning and higher academic achievement. Therefore, based on the results and the growing body of research that demonstrate their effectiveness in the classroom environment, it is recommended that educators become informed about the efficacious potential of metacognition and feedback in student learning.

For the students, reflective assessment provides an opportunity to take ownership of their learning on a regular basis. As suggested by Flavell (1979), by encouraging students to

reflect and "think about their thinking," they foster a skill set that transcends the classroom experience and benefits their long-term learning process. For the teachers, students' reflection informs their instruction to promote improved learning and to better know their students' thought processes.

There is an abundance of empirical evidence that supports the argument that reflective assessment positively impacts academic achievement (Bond, 2013; Bond & Ellis, 2013; Evans, 2009; Kramarski & Mevarech, 2003; Lew & Schmidt, 2011; Zan, 2000). Likewise, teacher feedback can positively and significantly impact student learning in terms of quality of homework, interest and motivation which lead to improved learning (Butler & Nisan, 1986; Nunez et al., 2015; Siewert, 2011). However, one could argue that many of these studies suffered from limitations in terms of the research design and data analysis. Additionally, further study is needed to provide empirical evidence that links both approaches, reflective assessment and feedback, as effective practices for improved learning. As such, it is vital that educational researchers continue to explore, research and refine the use of metacognitive practice linked with teacher feedback in the learning environment.

References

Arbuckle, J. L. (2009). *Amos 18.0 user's guide*. Crawfordville, FL: Amos Development Corporation.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York, NY: W.H. Freeman and Company.

Black P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5,* 7 – 74. http://dx.doi.org/10.1080/0969595980050102

Bond, J. (2003). *The effects of reflective assessment on student achievement* (Doctoral dissertation). Seattle Pacific University, WA.

Bond, J. B., & Ellis, A. K. (2013). The effects of metacognitive reflective assessment on fifth and sixth graders' mathematics achievement. *School Science and Mathematics, 113*, 227–234. http://dx.doi.org/10.1111/ssm.12021

Borich, G. D. (2014). *Effective teaching methods* (8th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Brookhart, S. (2008). *How to give effective feedback to your students.* Alexandria, VA: Association for Supervision and Curriculum Development.

Bruner, J. (1961). *The process of education*. Cambridge, MA: Harvard University Press.

Butler, R., & Nisan, M. (1986). Effects of no feedback, task-related comments, and grades on intrinsic motivation and performance. *Journal of Educational Psychology, 78,* 210-216. http://dx.doi.org/10.1037/0022-0663.78.3.210

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

Cho, H., & Abe, S. (2013). Is two-tailed testing for directional research hypotheses tests legitimate? *Journal of Business Research, 66,* 1261-1266. http://doi:10.1016/j.jbusres.2012.02.023

Costa, A. L. (2001). *Developing minds: A resource book for teaching thinking* (3rd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Dewey, J. (1910). *How we think.* Boston, MA: D.C. Heath & Co. publishers

Ellis, A. (2001). *Teaching, learning, and assessment together: The reflective classroom*. Larchmont, NY: Eye on Education, Inc.

Evans, L. (2009). *Reflective assessment and student achievement in high school English* (Doctoral dissertation). Seattle Pacific University, WA.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics.* Thousand Oaks: CA: SAGE Publication Ltd.

Flavell, J. H. (1977). *Cognitive development.* Englewood Cliffs, NJ: Prentice-Hall.

Flavell, J. H. (1979). Metacognition and cognitive monitor: A new area of cognitive development inquiry. *American Psychologist, 34,* 906 – 911. http://dx.doi.org/10.1037/0003-066X.34.10.906

Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational research: An introduction.* Boston, MA: A & B Publications.

Guskey, T. R., & Marzano, R. J. (2003). *Assessment as learning: Using classroom assessment to maximize student learning.* Thousand Oaks, CA: Corwin Press, Inc.

Guth, G. J. A., Holtzman, D. J., Schneider, S. A., Carlos, L., Smith, J. R., Hayward, G. C., & Calvo, N. (1999). Impact of standards-based accountability systems. In *Evaluation of California's standards based accountability system* (chapter 10). Retrieved from http://www.wested.org/online_pubs/accountability/SBA_chapter_10.pdf

Hall, C., Kennedy, B. J., Bass, D. L., & Wiggins, M. (2012). *Geometry Common Core.* Boston, MA: Pearson Prentice Hall.

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning.* London, UK: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. http://dx.doi.org/10.3102/003465430298487

Irons, A. (2008). *Enhancing learning through formative assessment and feedback.* New York, NY: Routledge.

Joyce, B., Weil, M., & Calhoun, E. (2009). *Models of teaching* (8th ed.). Boston, MA: Pearson Education, Inc.

Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal, 40*, 281-310. http://dx.doi.org/10.3102/00028312040001281

Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58,* 79–97. http://dx.doi.org/10.3102/00346543058001079

Leung, D. P., & Kember, D. (2003). The relationship between approaches to learning and reflection upon practice. *Educational Psychology: An International Journal of Experimental Educational Psychology, 23(1),* 61-71.

Lew, M. D. N., & Schmidt, H. G. (2011). Self-reflection and academic performance: Is there a relationship? *Advance in Health Science Education, 16*, 529–545. http://dx.doi.org/10.1007/s10459-011-9298-z

Marzano, R. J., Boogren, T., Heflebower, T., Kanold-McIntyre, J., & Pickering, D. (2012). *Becoming a reflective teacher.* Bloomington, IN: Marzano Research Laboratory.

Mevarech, Z. R., & Kramarski, B. (1997). IMPROVE: A multidimensional method for teaching mathematics in heterogeneous classrooms. *American Educational Research Journal, 34*, 365–394. http://dx.doi.org/10.3102/00028312034002365

Nunez, J. C., Suarez, N., Cerezo, R., Gonzalez-Pienda, J. A., Rosario, P., Mourao, R., & Valle, A. (2013). Homework and academic achievement across Spanish compulsory education. *Educational Psychology, 35,* 726-746. http://dx.doi.org/10.1080/01443410.2013.817537

Nunez, J. C., Suarez, N., Rosario, P., Vallejo, G., Cerezo, R., & Valle, A. (2015). Teacher's feedback on homework, homework-related behaviors, and academic achievement. *The Journal of Educational Research, 108,* 204-216. http://dx.doi.org/10.1080/00220671.2013.878298

Popham, W. J. (2014). *Classroom assessment: What teachers need to know* (7th ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Ramsden, P. (2003). *Learning to teach in higher education* (2nd ed.). London, UK: Routledge.

Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice (5)*, 77 – 84. http://dx.doi.org/10.1080/0969595980050104

Schoenfeld, A. H. (1987). *Cognitive science and mathematics education.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Siewert, L. (2011). The effects of written teacher feedback on the academic achievement of fifth-grade students with learning challenges. *Preventing School Failure, 55,* 17-27. http://dx.doi.org/10.1080/10459880903286771

Stiggins, R. J. (1996). *Student-centered classroom assessment.* Upper Saddle River, NJ: Merrill Prentice Hall.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological process.* Cambridge, MA: Harvard University Press.

Wiliam, D. (2012). Feedback: Part of a system. *Educational Leadership, 70,* 31-34.

Zan, R. (2000). A metacognitive intervention in mathematics at university level. *International Journal of Mathematical Education in Science and Technology, 31*(1), 143–150. http://dx.doi.org/10.1080/002073900287462