12-2015

Improving the accuracy of middle school students' self-assessment, peer assessment, and mathematics achievement.

Elizabeth Popelka
University of Louisville

IMPROVING THE ACCURACY OF MIDDLE SCHOOL STUDENTS' SELF-ASSESSMENT, PEER ASSESSMENT, AND MATHEMATICS ACHIEVEMENT

By

Elizabeth Popelka
B.A., University of Louisville, 1999
M.A.T., University of Louisville, 2006

A Dissertation
Submitted to the Faculty of the
College of Education and Human Development of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy in Curriculum and Instruction

Department of Teaching and Learning
University of Louisville
Louisville, Kentucky

December 2015

IMPROVING THE ACCURACY OF MIDDLE SCHOOL STUDENTS' SELF-ASSESSMENT, PEER ASSESSMENT, AND MATHEMATICS ACHIEVEMENT

By

Elizabeth Popelka
B.A., University of Louisville, 1999
M.A.T., University of Louisville, 2006

A Dissertation Approved on

November 23, 2015

by the following Dissertation Committee:

____

Dissertation Director
Karen S. Karp

____

Elizabeth Todd Brown

____

Maggie B. McGatha

____

Jeffrey C. Valentine

____

Robert N. Ronau

DEDICATION

In loving memory of my mother,

Joy Mae Popelka,

the first teacher I ever knew, who instilled in me an appreciation for education and the wonderful places it could take me.

ACKNOWLEDGMENTS

I began this journey to my doctorate because I love being a teacher and welcome every opportunity to enhance my practice. Having learned so much from my coursework and gained so much from my research, and before I can finally give back to the field and make my own contribution, I am thrilled to recognize those who made my academic experience so meaningful.

I am forever grateful to the members of my dissertation committee for their contributions to this study and to my overall growth as an educator. As my Rank 1 program advisor and professor during my early teaching years, Dr. Maggie McGatha laid an appreciation of formative assessment that would become the foundation for this dissertation. Over my coursework and our collaboration on multiple projects, Dr. Elizabeth Todd Brown has been an amazing teacher and mentor, delivering positive—yet honest—pieces of advice that have guided me through my academic experience. Before studying with Dr. Jeffrey Valentine, I had been lucky to muddle through my statistics classes. I appreciate so much his patience with my lack of background knowledge, and I am awed by how much he has taught me about applying statistics to realistic situations. For everything I know about research design and advanced statistical analysis, I must credit Dr. Robert Ronau. The number of hours he has logged in helping me construct and fully grasp the implications of this study is so high I had to quit estimating. Without his efforts, this dissertation would not have been possible.

Lastly, I would like to thank my program advisor and dissertation chair, Dr. Karen

Karp. Because her reputation precedes her, I was flattered when she advised me during my Rank 1 program to pursue a Ph.D. Because of her famously high standards, I was elated when she agreed to be my advisor and dissertation chair. The last several years have produced a number of professional and personal ups and downs that drew out the dissertation process for longer than either of us had planned or would have liked. Because she is not only an incredible teacher and mentor but also a dear friend, Dr. Karp was incredibly sympathetic to and supportive of me even as the length and duration of my dissertation became a source of hardship for her. I am a better student, researcher, teacher, and person, because of Dr. Karp. I am honored to join her impressive list of other "Ph.D. children."

This project truly took a village, and I was lucky to have the most nurturing one I could have hoped for: my colleagues, friends, and "family" at Westport Middle School. I am grateful for the school administrators who provided me access to classes and, most especially, to the fellow teachers who helped validate my rubric validation, examined assessments for validity, relinquished their students to me, and acted as a sounding board when I needed to talk out ideas and frustrations. The statistical analysis required for this study was, at first, beyond the scope of my comprehension, and certain university faculty went over and beyond to help me understand it and carry it out with confidence and ease. I am forever indebted to the following people, all of whom contributed directly to this dissertation: Dr. Christopher Rakes, Dr. Staci Eddleman, Jodie Zeller, Sagan Massey, Kristyn Williams, James Daniels, Lauren Thomas, Dave Thomas, Meagan Anderson, Heather Weis, Jennifer Napier, Emily Hixon, Stacey Hubbard, Torri Martin, Sharon Klump, Laura Viergutz, Jonathan Lippman, Sarah Yost, Melissa Shearon, Dr. Elizabeth

Best, Judith Reader, and Mark Dumouchelle.

Talking of strategies to survive the dissertation process, a wise woman, Dr. Kate Grindon, urged, “Find your tribe.” She was totally right: Nobody can understand this process like someone who has gone through it. I could not ask for better friends and Ph.D. sisters than Victoria Miller Bennett, Shannon Stone, and Leah Dix White. And neither could I ask for a more invaluable “big” Ph.D. sister, Dr. Sarah Bush.

Finally, all the pages in the world cannot supply adequate words to thank my amazing husband, Ray Brown. In order for me to finish this seemingly never-ending process, he picked up an unbelievable amount of family responsibility, displayed infinite patience and support for my goals, sacrificed countless hours of quality couple and family time, and continued to love me even though, these last couple of years, I’ve had many moments when I was very unlovable. He is my biggest booster and my best friend. I cannot imagine this journey, academic and otherwise, without him.

ABSTRACT

IMPROVING THE ACCURACY OF MIDDLE SCHOOL STUDENTS' SELF-ASSESSMENT, PEER ASSESSMENT, AND MATHEMATICS ACHIEVEMENT

Elizabeth Popelka

November 23, 2015

Despite the documented advantages of formative assessment (FA) strategies in elevating student achievement, much of the relevant research is dated and undermined by questionable design and inattention to K-12 settings. In order to fill these voids, this quantitative study tested the effect of a self- and peer-assessment-training instructional sequence, developed with recommendations from past research and employing explicitly described assessment measures and criteria, on middle school students' assessment accuracy and subsequent mathematics achievement. The researcher hypothesized a correlation between measurable growth in assessment accuracy and gains in achievement, as well as a reciprocal effect of self- and peer-assessment ability.

The subjects for this study were drawn from the population of 7th-grade students at a low-achieving urban middle school in a metropolitan area in the Midwest. The students were convenience- and purposive-selected and assigned to treatment and control groups; the treatment group contained 39 students and the control group contained 41 students. Data were collected before and after 10 hour-long self- and peer-assessment training sessions, which were conducted with the treatment group over a 3-week period.

This study had an independent variable with two groups, treatment and no treatment, and three dependent variables, achievement, students' ability to accurately self-assess, and students' ability to accurately peer-assess. All variables were assessed numerically and analyzed using a multivariate statistical procedure. Statistical tests revealed a positive effect of peer- and self-assessment training on students' mathematical achievement and ability to accurately self-assess. The intervention did not increase the accuracy of students' peer assessments, contradicting the foundational conjecture that self- and peer-assessment ability would rise commensurately. Nevertheless, the gains in self-assessment accuracy and student achievement produced by the intervention indicate that teachers should adopt assessment training in classrooms. Ultimately, this study contributes a clear and demonstrably effective instructional sequence; exemplifies successful strategy implementation; and freshly affirms the connection between student assessment practices and achievement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE PAGE

CHAPTER I: INTRODUCTION

Problem Statement

"The future well-being of our nation and people depends not just on how well we educate our children generally, but on how well we educate them in mathematics and science specifically" (U.S. Department of Education [USED], 2000, p. 4), declared the United States Department of Education at the turn of the millennium. Unfortunately, data suggest that many U.S. students have not been achieving at levels in mathematics adequate to realizing this ambition. According to the 2013 National Assessment of Educational Progress (NAEP) data, 58% of all U.S. fourth graders and 64% of all U.S. eighth graders tested at either basic or below-basic levels of mathematics achievement. Although historic NAEP data reflect growth in U.S. students' mathematics proficiency over the last two decades (USED, 2013), current evidence nevertheless indicates underperformance. Given the relationship between mathematics achievement and the ability of U.S. citizens to compete in an increasingly global economy (National Council of Teachers of Mathematics [NCTM], 2013a), the "middle of the pack" (National Science and Technology Council [NSTC], 2013, pp. vi-vii) performance by U.S. K-12 students in a recent international comparison study are troubling to educators and national leaders alike.

Existing Literature

Development of Common Core State Standards

In response to concerns that U.S. students were academically underperforming relative to students in other nations, in 2009, the National Governors' Association Center for Best Practices and the Council of Chief State School Officers commissioned the development of the Common Core State Standards (CCSS), a set of academic standards for K-12 English language arts and mathematics (National Governors Association Center for Best Practices & Council of Chief State School Officers [NGACBP & CCSSO], 2010). The CCSS were crafted to meet a number of goals: assessments and instruction based on a consistent and clear set of knowledge and skills (Conley, 2011); a narrower and more-in-depth curriculum, like that found in higher-achieving countries (Porter, McMaken, Hwang, & Yang, 2011); and a focus on instructional practices that prepare all students for college and the workforce (NGACBP & CCSSO, 2010, About the Common Core State Standards section).

The underpinning of the Common Core State Standards for Mathematics (CCSSM) is years of research on mathematics education practices from the National Research Council and investigation into effective state and international educational practices and standards (Mathematics Common Core Coalition [MCCC], n.d.). The CCSSM represent a shift toward "greater focus on fewer topics, coherence: linking topics and thinking across grades, and rigor: pursue conceptual understanding, procedural skills and fluency, and application with equal intensity" (NGACBP and CCSSO, 2010, Key Shifts in Mathematics section). These components of the CCSSM are intended to not only "build on the best of high-quality math standards from states across the country," but "also draw on the most important international models for mathematical practice" and embody "research and input from numerous sources, including state departments of education, scholars, assessment developers, professional organizations, educators, parents and students, and members of the public" (NGACBP & CCSSO, 2010, Mathematics Standards section, para. 2). The CCSSM have been supported by the NCTM, whose past president Linda Gojak declared, "The Common Core State Standards call national attention to what is important for students to know and be able do, not only to be career and college ready, but also to be quantitatively literate citizens—they, after all, are the ones who will make decisions that shape the future" (NCTM, 2013c, para. 2).

Standards for Mathematical Practice

Influenced by the NCTM process standards (2000) and the National Research Council's *Adding It Up* (2001), the CCSSM Standards for Mathematical Practice (SMP) are a list of habits that proficient mathematicians exhibit and that should be cultivated within every mathematics student (NGACBP & CCSSO, 2010, Standards for Mathematical Practice section). For example, SMP 3—"Construct viable arguments and critique the reasoning of others" (NGACBP & CCSSO, 2010, Standards for Mathematical Practice section, para. 4)—derives from the belief that argumentation and critique help a student "reflect on his or her own reasoning and understanding" (Halani, Davis, & Roh, 2013, p. 1). Emphasis on this type of critical thinking, it is contended, results in higher student engagement and academic achievement (NSTC, 2013).

Impact of Assessment

While the aforementioned changes to the curriculum across the US may prove to drive student achievement, it has been argued that the most important factors shaping students' learning occur in the classroom (Black & Wiliam, 1998b; Noonan & Duncan, 2005). Extending this is the claim that no facet of the classroom experience can shape students' education and their subsequent futures more than a teacher's assessment practices. Stiggins and Chappuis (2005) articulated this notion:

> From their very earliest school experiences, our students draw life-shaping conclusions about themselves as learners on the basis of the information we provide to them as a result of their teachers' classroom assessments. As that evidence accumulates over time, they decide if they are capable of succeeding or not. They decide whether the learning is worth the commitment it will take to attain it. They decide if they should have confidence in themselves as learners and in their teachers—that is, whether to risk investing in the schooling experience. These decisions are crucial to their academic well-being. Depending on how they decide, their teachers may or may not be able to influence their learning lives. (p. 11)

One currently inevitable aspect of the classroom experience is high-stakes testing (Black & Wiliam, 2009). Previous research attests that high-stakes testing can greatly impact students' motivation and self-esteem (Nicol & Macfarlane-Dick, 2006), among other critical student characteristics that have been linked to achievement (Clark, 2012; Stiggins & Chappuis, 2005), either positively or negatively, depending on the quality and content of the assessment (Stiggins & Chappuis, 2005). Because of the power that teachers' assessment practices can wield over their students' potential academic success, it has been remarked, teachers must carefully consider the choices they do have about assessment and employ strategies that most benefit the greatest number of students (Heritage, 2007).

Formative Assessment

Ample research has demonstrated that student achievement can be raised through formative assessment (FA) (Black & Wiliam, 1998a; Crooks, 1988; Leahy & Wiliam, 2012; Stiggins, 2002; Wiliam, 2007b). Underpinned by post-structuralism, FA is a set of strategies united by a belief in employing assessment as a tool to improve learning (Wiliam, 2007b). This notion contradicts past common conceptions of assessment as separate from, or even a detraction from, instruction and as a way to measure the content of what students have learned (Heritage, 2007). Effective formative assessments are embedded within instruction (Suurtamm, Koch, & Arden, 2010), and the results they yield lead to needed instructional adjustments while the learning is actually taking place (Leahy, Lyon, Thompson, & Wiliam, 2005) and while students still have the chance to enhance their understanding and the quality of their work (Schoenfeld, 2015). Scholars have noted that the use of the term "formative" to denote this kind of assessment has engendered misunderstanding, because "formative" has sometimes been used to refer to any assessment given before a major high-stakes test; examples include the multitude of interim and benchmark assessments that have been purchased in increasing numbers by districts, assessments dismissed by Heritage (2007) as nothing more than "early warning summative" assessments (p. 140). According to FA theory, an assessment is formative when its results are used to direct students' future learning activities, irrespective of the assessment's format or timing. If teachers do not base subsequent instructional decisions on these results, the assessment is not formative (Davis & McGowen, 2007).

Benefits of an FA approach. Black and Wiliam (1998a) reviewed approximately 250 previous studies to examine the connection between the use of FA strategies and student learning gains. The authors measured said gains by comparing average student test score increases that followed implementation of FA techniques with scores that did not. This analysis indicated that when a FA approach was used, learning gains were achieved across age groups, content areas, and even several countries, and that gaps between low and high achievers were closed. Subsequent studies have corroborated Black and Wiliam's findings and engendered a consensus among scholars that FA practices do enhance student outcomes (Andrade & Valtcheva, 2009; Black & Wiliam, 1998b; Brown & Hirschfeld, 2007; Davis & McGowen, 2007; Heritage, 2007; Heritage & Heritage, 2013; Leahy et al., 2005; Nicol & Macfarlane-Dick, 2006; Ross, Hogaboam-Gray, & Rolheiser, 2002; Suurtamm et al., 2010; Thomas, Martin, & Pleasants, 2011; Wiliam, 2007a).

Student role in the FA process. The above description of FA as a tool for subsequent instructional decisions might suggest that FA gains are simply a product of a set of actions that teachers implement. In fact, the strides stem from actions encouraged among and taken by students, namely, their expanded participation in assessment processes, a core tenet of FA processes (Black & Wiliam, 2009; Clark, 2012; Heritage, 2007; Noonan & Duncan, 2005; Stiggins & Chappuis, 2005). To maximize the potential of FA, the responsibility for learning must be shared by the teacher with students (Black & Wiliam,1998b). Touting the teacher-student assessment collaboration, Thomas et al. (2011) stated that "[w]hen teachers share with their students the process of assessment—giving up control, sharing power and leading students to take on the authority to assess themselves—the professional judgment of both is enhanced" (p. 1). This notion was reinforced by Voogt and Kasurien (2005), who wrote, "Formative assessment may consist of hard data, but more often and more importantly of 'tacit knowledge', i.e.

knowledge that both the teacher and student obtain through discussion, reflection and experience" (p. 8).

Feedback in FA. At the heart of the formative process of shared responsibility, experiences, and dialogue is feedback (Clark, 2012). Underscoring the centrality of feedback in mathematics education, the NCTM's (2013a) five-paragraph-long position statement on FA mentions the important role of teacher- and student-provided feedback seven times. In the statement, it is observed that teacher-provided feedback helps students make accurate judgments about their current progress and develop plans for improvements to subsequent work; it is also remarked that student-provided feedback offers teachers insights into students' thinking, so that they may adjust subsequent instruction to better meet students' needs. Helpful feedback is not only produced during teacher-student interactions; students can be resources for themselves and each other when they generate feedback while engaged in self- and peer assessment (Black & Wiliam, 1998a; Clark, 2012; Sluijsmans, Dochy, & Moerkerke, 1999; Sung, Chang, Chiou, & Hou, 2005), two strategies indispensable to any FA-based curriculum (Leahy et al., 2005).

Self-assessment and peer assessment. Countering the notion of self- and peer assessment as totally separate processes with their own respective benefits, Noonan and Duncan (2005) argued that they are in fact complementary and overlap. Central to both is comparing a work product or performance against a pre-established set of criteria or goals and determining the degree to which the product meets the criteria. The chief differences lie in the audience and the assessor. During self-assessment, the student engages in an analysis of his or her own work, while during peer assessment, the student analyzes the work of a classmate (Clark, 2012; Liu & Carless, 2006).

Because the formative nature of students' assessments discourages grade assignment, it likewise instills in students the view of assessment as a learning experience, rather than an evaluative one (Nicol & Macfarlane-Dick, 2006), through what is called a "mastery" focus (Guskey, 2003, p. 10). With a mastery focus, students are urged to treat assessment as an opportunity to gain information about present performance as a pathway toward improved future performance, not merely to "pass"; thus, they are less likely to incur crippling feelings of failure (Black & Wiliam, 2009). It has been noted that students' view of what matters most in their education is shaped by the assessments they take (Ellis & Folley, 2010). As they engage in mastery-focused assessments over time, the perception of assessments as a learning experience can augment their intrinsic motivation (Clark, 2012; Noonan & Duncan, 2005; Ross, 2006) and perseverance (Clark, 2012; Noonan & Duncan, 2005)—characteristics that, as noted, have a positive correlation with achievement (Clark, 2012).

The FA-oriented mastery focus of students' assessments promotes learning by having students pursue goals of ever-improving work and performance (Ross et al., 2002; Wiliam, 2007a). Recipients of peer-assessment feedback often profit from a more objective set of eyes that can, typically, identify ways to improve work more successfully than the author can (Leahy et al., 2005). These opportunities for improvement can often be more easily pinpointed by peers and expressed in more-accessible terms than those used by their teachers (Nicol & Macfarlane-Dick, 2006). When providing peer feedback, students practice and enhance their evaluation skills without the negative emotions that can, and frequently do, arise from analyzing their own products (Wiliam, 2007a). With better-honed evaluation skills, students grow their capacity for self-regulation, which Eva and Regehr (2011) noted includes the process of identifying and addressing opportunities for improvement and Black and Wiliam (1998a) cited as another skill directly related to achievement.

When teachers make students full partners in the assessment process, they impress upon students that they are valued and trusted and that their input matters (Heritage, 2010b; Stiggins & Chappuis, 2005). The supportive environment thereby created contributes to a number of factors that directly feed achievement (Clark, 2012). When students believe they have an important role to play, they experience increases in their sense of self-efficacy (Clark, 2012), their sense of control over their own academic progress (Butler & Winne, 1995), their motivation (Liu & Carless, 2006), and their responsibility in the learning process (Nicol & Macfarlane-Dick, 2006). In addition, through increased dialogue, students gain valuable skills in effective collaboration and cooperation (Heritage, 2007).

The fruits of student assessment—namely, elevated achievement and gains in the qualities that promote achievement—are firmly established in the literature. Benefits have been found across content, including mathematics, (Andrade & Valtcheva, 2009; Leahy et al., 2005; Ross, 2006; Ross et al., 2002; Warner, Chen, & Andrade, 2012), writing (Ross, 2006), science (Andrade & Valtcheva, 2009), social studies (Andrade & Valtcheva, 2009; Ross, 2006), and in non-academic areas, such as increased positive group interactions and reduced disruptive behavior (Ross, 2006). Benefits also have been found across grade levels, including elementary and middle school (Andrade &

Valtcheva, 2009; Ross et al., 2002; Warner et al., 2012), high school (Andrade & Valtcheva, 2009; Brown & Hirschfeld, 2007; Leahy et al., 2005; Onion & Javaheri, 2011; Warner et al., 2012) and college (Andrade & Valtcheva, 2009; Nicol & Macfarlane-Dick, 2006; Thomas et al., 2011).

Accuracy of students' self- and peer assessments. Despite research enumerating the virtues of many FA strategies, teachers often decline to incorporate them into their regular practice, citing what they perceive to be insurmountable obstacles, such as time constraints or lack of training in assessment (Guskey, 2003). More significantly, teachers do not have students undertake self- and peer assessment (Noonan & Duncan, 2005) out of the belief that students cannot or will not accurately assess themselves (Andrade & Du, 2007; Bryant & Carless, 2010). These beliefs are, to a degree, supported by research. The literature, which includes meta-analyses and individual studies ranging over a variety of contexts, appears to reveal low, if any, levels of agreement between teachers' and students' assessments (Andrade & Du, 2007; Boud & Falchikov, 1989; Falchikov & Boud, 1989; Fox & Dinur, 1988; Narciss, 2008; Nulty, 2011; Pakaslahti & Keltikangas-Järvinen, 2000; Ross, 2006; Topping, 1998). Even many of the studies that show consistency between teachers' and students' assessments used questionable research designs, particularly, the lack of a clear description of measurement scales and the introduction of subjective criteria into the assessment (Boud & Falchikov, 1989; Falchikov & Boud, 1989; Topping, 1998). In addition, most of the research that has produced relatively higher levels of accurate self- and peer assessment, including that obtained subsequent to successful implementation of interventions, has taken place in a higher-education setting (Kulkarni et al., 2013; Lawson et al., 2012; Wagner, Suh, &

Cruz, 2011). K-12 teachers, therefore, may doubt that they can replicate the results in their own classrooms.

Statement of Purpose

Given teachers' fears about the potential inaccuracy of and resultant resistance to implementing student self- and peer-assessment practices and the dearth of research about interventions for improving middle school students' self- and peer-assessment accuracy, this study conducted a data analysis to determine whether an intervention, developed with recommendations from previous research, positively affected students' mathematics achievement and their ability to accurately self- and peer-assess.

Research Questions

The study addresses the following research questions:

1) What is the effect of self- and peer-assessment training of middle school students on their mathematics achievement?

2) What is the effect of self- and peer-assessment training of middle school students on their ability to accurately assess themselves?

3) What is the effect of self- and peer-assessment training of middle school students on their ability to accurately assess their peers?

Significance of the Study

There is an abundance of previous research describing the benefits of and ways to incorporate self- and peer assessment. However, research, and recent research especially, on measuring or improving the accuracy of students' self- and peer assessment is sparse, and little is based in a middle school setting. Moreover, most of the available research may be faulted for lacking clear assessment measures and criteria. This study used explicitly described assessment measures and criteria both to quantitatively determine the effect of a self- and peer assessment instructional sequence on middle school students' ability to self-assess and peer assess and to improve their mathematics achievement.

Delimitations

This study had several delimitations. The study was conducted in late April and early May 2015 in an urban Midwestern middle school. The sample consisted of students enrolled in the mathematics classes of one of two participating seventh-grade mathematics teachers. The intervention was delivered over 3 weeks in 10 hour-long sessions during the school day.

Assumption

The main assumption underlying this study is that teachers are accurate in their assessments of students' level of success with task completion. Therefore, the students' assessment accuracy will be calculated from the discrepancy between their assessments and my own, as the teacher overseeing the intervention.

Summary

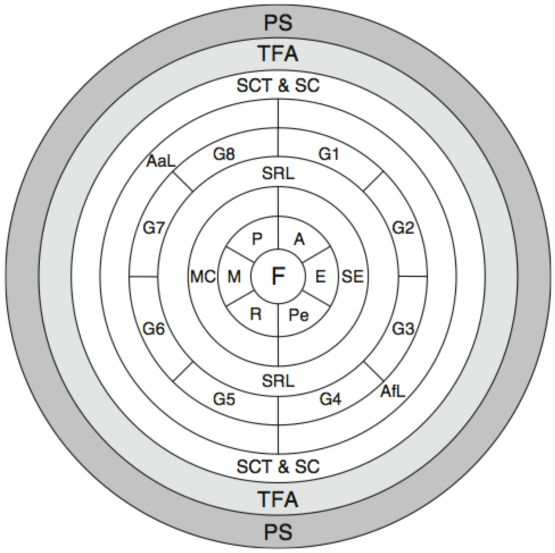
Recently, U.S. students' mediocre mathematics performance relative to that of many international peers has contributed to a movement toward reform in the national mathematics curriculum. However, research has shown that the most effective agents of positive academic achievement are the teacher-implemented strategies in individual classrooms. Certain assessment strategies have demonstrated great potential to enhance learning, including student self-assessment and peer-assessment. However, despite the benefits, teachers are frequently hesitant to incorporate them, citing doubts that students can accurately assess themselves and their peers. Although some research does attest that students can accurately and effectively self-assess, especially after they are explicitly taught how, much of it is dated, of questionable quality, or does not take place in a middle school setting. This study seeks to fill these gaps by testing the effect of an intervention, developed from recommendations of past research, on 7th-grade-students' achievement and self- and peer-assessment accuracy.

CHAPTER II: LITERATURE REVIEW AND THEORETICAL FRAMEWORK

The following chapter will provide the theoretical background and components that contributed to the development of this study. The first three sections will describe each component of the study intervention: Clark's Theory of Formative Assessment (TFA), self-assessment, and peer assessment. In the final section, I will present the background behind the intervention's theorized effect on the three dependent variables—namely, student ability to self-assess, student ability to peer-assess, and students' mathematics achievement—as well as interaction that could occur among the variables subsequent to the intervention.

Intervention Component 1: Clark's Theory of Formative Assessment

Clark's TFA (2012) provides most of the theoretical background for the proposed study intervention. In his model, Clark synthesized a foundation for the development of a TFA, including key objectives and beliefs. The following framework (p. 207) illustrates Clark's model and its grounding in post-structuralism:



PS - post-structuralism, **TFA -** theory of formative assessment, **SCT & SC -** socio-cognitive theory and sociocultural theories, **AaL -** assessment as learning, **AfL** - assessment for learning, **G1 - G8** formative goals, **SRL** - self-regulated learning, **MC** - metacognition, **SE -** self-efficacy, **P -** planning, **M -** monitoring, **R -** reflecting, **A** - ambition, **E -** effort, **Pe -** persistence, **F-** feedback

Figure 2.1. Clark's TFA, shown to be grounded in post-structuralism and to initiate a comprehensive process of self-regulation, in cross-section.

In the outermost ring in Clark's framework diagram, formative assessment theory is encompassed within post-structuralism (PS). According to Peters (1999), PS is a multifaceted philosophy that rejects the notion that scientism can be widely applied to research involving human beings and emphasizes the influence of a person's subjectivity when he or she is making interpretations. PS shares with constructivist theory (Vidmar, 2011) the idea that students are not simple receptacles but active constructors of knowledge who assimilate previous learning with new information and stimuli (Zimmerman, 2001).

Formative Assessment

Concepts under the PS umbrella include the importance of individual differences and emphasis on learners' construction of new knowledge, both major themes of formative assessment (FA) (Clark, 2012). As opposed to a process separate from teaching, FA "needs to be intentionally and systematically integrated into classroom instruction at every grade level" (National Council of Supervisors of Mathematics & Association of Mathematics Teacher Educators [NCSM & AMTE], 2014). Embedded thus within instruction (Suurtamm et al., 2010), FA is a process that generates and uses assessment results to tailor subsequent learning goals and experiences to the unique needs of students (Leahy et al., 2005).

Yet a misapprehension exists among some educators that an assessment's "formative" quality is determined by its particular characteristics, such as timing or format, rather than by how its data are collected and used (Shepard, 2009). For example, in recent years, assessment vendors have marketed to schools certain assessments mislabeled "formative," with the claim that they diagnose students' knowledge and understanding gaps and improve subsequent performance on high-stakes testing (Perie, Marion, & Gong, 2009). Many educators succumb to the misinformation, in part, because of the timing of the assessments' administration, before summative assessments (Heritage, 2007). In reality, many of these assessments are interim assessments, "assessments administered during instruction to evaluate students' knowledge and skills relative to a specific set of academic goals in order to inform policymaker or educator decisions at the classroom, school, or district level" (Perie et al., 2009, p. 6). While it can and has been argued that interim assessments fill a need in education, many are not structured to provide information that is detailed enough to inform subsequent instruction (Shepard, 2009) and, thus, are not formative. While a formative assessment can appear in a wide variety of formats, its formative character is derived from the way information gleaned from the assessment is used, not from the assessment format itself; if subsequent learning and instructional decisions are not based on assessment information, the assessment cannot be called formative (Black & Wiliam, 1998a; Black & Wiliam, 2009; Davis & McGowen, 2007; NCSM & AMTE, 2014). The spontaneous nature of "on the fly assessment" formats (Heritage, 2007, p. 141), usually observations (Stepanek & Jarrett, 1997) of students' dialogue, actions, and facial expressions (Leahy et al., 2005), typically qualifies these types of assessment as formative. In addition, "planned-for interaction" formats (Heritage, 2007, p. 141), such as white boards, student response clickers (Wallace, 2003), questioning (Black & Wiliam, 2009; Clark, 2012), and student interviews (Stepanek & Jarrett, 1997), are usually considered formative, because they are more-frequent, shorter assessments that allow for quick data collection and subsequent changes to instruction (Black & Wiliam, 1998a). However, some assessment types once categorized as summative are now being used formatively, such as portfolios (Stepanek & Jarrett, 1997) and even tests (Black & Wiliam, 2009).

Benefits of formative assessment. When assessment results are used to guide future instructional decisions, students demonstrate dramatic learning-rate increases, ranging from as much as 50% to 100% (Leahy & Wiliam, 2012), over the short, medium (Wiliam, 2007b), and long term (Leahy et al., 2005). These gains are reflected in improved classroom summative test scores (Crooks, 1988; Stiggins, 2002) and standardized test scores (Stiggins & Chappuis, 2005), across a variety of age groups, countries, and content areas (Black & Wiliam, 1998a; Leahy et al., 2005). Some may regard as especially significant the fact that FA techniques have been shown to raise achievement levels for all students while lessening the gap between low achievers and others (Black & Wiliam, 1998a).

Assessment for learning and assessment as learning. Laying the groundwork for a more specific list of principles and strategies to be employed during the implementation of FA, Clark's (2012) framework, using two adjoining sections within his TFA, illustrates two primary goals: assessment for learning (AfL) and assessment as learning (AaL). These two concepts were developed, like FA, in response to the increase in assessment of learning (AoL), which is used to determine how much students have learned. According to critics, AoL produces scores too late and with insufficient information to be useful to plan subsequent student instruction (Clark, 2012; Guskey, 2003; Heritage, 2007). By contrast, AfL and AaL are designed to impact subsequent instructional decisions and enhance students' learning outcomes (Clark, 2012; Stiggins, 2002).

Assessment for learning. The description of AfL might lead one to believe, erroneously, that AfL and FA are synonymous (Clark, 2012). While FA does have as one of many goals increased student participation in the assessment process (Thomas et al., 2011), a teacher can obtain assessment data and plan subsequent instruction—in other words, can use FA—with no student action outside of taking the assessment. On the other hand, AfL activities—such as providing students with assessment criteria and corrective feedback—although teacher driven, require students' participation (Stiggins, 2002). As Stiggins and Chappuis (2005) framed it, AfL provides both "teachers and students with information they need along the way, during the learning process, to make decisions that will bring about more learning. In this side of the assessment house, an effective communication system provides regular diagnostic information to the teacher and frequent descriptive feedback to the learner" (p. 6). Simply put, one may use FA without using AfL, but one cannot use AfL without applying FA. For example, a teacher might give a quiz to students and then use the results to plan subsequent remediation lessons on a topic, if needed. In this case, the teacher is using FA but not AfL. If the teacher conferences with a student after the same quiz and provides evidence-based feedback regarding ways to improve performance over the quiz topic, the teacher is using AfL, a specific instantiation of FA.

Assessment as learning. Although AfL embraces a student-centered approach (Noonan & Duncan, 2005), the focus is on the practices of the teacher, such as questioning and providing feedback; in contrast, AaL is based on activities that make students full partners in assessment (Clark 2012). The literature has established that, in order to most effectively promote students' self-regulation, teachers must allow students to be partners in the assessment process (Black & Wiliam, 2009; Clark, 2011; Heritage, 2007; Noonan & Duncan, 2005; Stiggins & Chappuis, 2005). A classroom with a focus on AaL is one in which "pupils and staff set learning goals, share learning intentions and success criteria, and evaluate their learning through dialogue and self and peer assessment" (Clark, 2012, p. 4). When engaged in peer assessment, students "provide either feedback or grades (or both) to their peers on a product, process, or performance, based on the criteria of excellence for that product or event" (Thomas et al., 2011, p. 3); when engaged in self-assessment, students evaluate their own current perceived level of performance with pre-established criteria for success (Lawson et al., 2012).

Self-regulated learning. Research-based support for the student-centered FA goals of AfL and AaL (Clark, 2011; Noonan & Duncan, 2005; Stiggins & Chappuis, 2005; Thomas et al., 2011) stems from the positive impact these practices have on the development of students' self-regulation (Black & Wiliam, 2009; Clark, 2012; Nicol & Macfarlane-Dick, 2006), the "tendency to monitor and manage one's own learning" (Andrade & Valtcheva, 2009, p. 13). In classrooms where self-regulated learning (SRL) is cultivated, students in grades as early as kindergarten to third grade have been observed independently conducting research, monitoring and evaluating their progress in the writing process, and strategically seeking support from peers (Perry, VandeKamp, Mercer, & Nordby, 2010). Learners are conscious of their use of these dynamic self-regulation strategies (Butler & Winne, 1995), "those processes, internal and/or transactional, that enable an individual to guide his/her goal-directed activities over time and across changing circumstances" (Clark, 2012, p. 216).

Students who engage in higher levels of self-regulation, those who "self-generate thoughts, feelings, and actions to attain their learning goals" (Zimmerman, 2001, p. 5), are generally higher-achieving than those who do not (Andrade & Valtcheva, 2009; Brown & Hirschfeld, 2007; Nicol & Macfarlane-Dick, 2006). Evidence of the correlation between students' self-regulation and achievement is so pronounced that self-regulation has been called "a pivot upon which students' achievement turns" (Butler & Winne, 1995, p. 245). For example, in one study of 10th-grade students, 93% of the students could be accurately categorized as high- or low-achieving, based on the results of interviews that ascertained their degree of self-regulation-strategy use (Zimmerman & Pons, 1986). Students who are better self-regulators also demonstrate heightened levels of other qualities correlated with high achievement. Among these are confidence, independence (Nicol & Macfarlane-Dick, 2006), willingness to collaborate (Heritage, 2007), perseverance, responsiveness to feedback (Clark, 2012), a larger repertoire of strategies to apply to various tasks (Black & Wiliam, 2009), positive attitudes toward learning, and openness to tackling challenges, a quality found even among relatively low-achieving self-regulating students (Perry et al., 2010).

Beyond their proportionally higher achievement, highly self-regulating students are also better equipped to apply previously learned skills and knowledge to new situations (Clark, 2012). Such a trait is critical for success in the current world of proliferating innovation, where job skills "are no longer passed down from one generation to the next" and "where it is not possible to establish which type of knowledge is needed in the next 5 or 10 years let alone a lifetime" (Clark, 2012, p. 233). By adopting teaching and learning strategies that advance students' capacities for self-regulation, educators are increasing the likelihood that they will be "lifelong learners" (Falchikov & Boud, 1989, p. 395) and positively impact their own future well-being.

Metacognition and self-efficacy. Encapsulated by self-regulated learning in Clark's (2012) TFA framework are the two components that operationalize self-regulation: metacognition and self-efficacy. Self-regulation, according to Clark, could not exist without either of these two components; therefore, as FA, through AfL and AaL practices, improves students' self-regulation (Black & Wiliam, 2009; Clark, 2012; Nicol & Macfarlane-Dick, 2006), AfL and AaL practices likewise enhance metacognition and self-efficacy (Clark, 2012). Metacognition may be defined as "the deliberate conscious control of cognitive activity" (Schunk, 2012, p. 286)—in other words, when someone intentionally thinks about their thinking—and comprises learners' engagement in goal-planning, monitoring, and reflection (Clark, 2012). Self-efficacy refers to individuals' "beliefs about their capabilities to exercise control over their own level of functioning and over events that affect their lives" (Butler & Winne, 1995, p. 256), and, according to Clark's (2012) framework, is composed of students' ambition, effort, and persistence.

Socio-cognitive theory and sociocultural theory. In Clark's (2012) framework diagram, AfL, AaL, and self-regulation are all enclosed within socio-cognitive theory and sociocultural theory. These theories, as the diagram manifests, inform and shape AfL and AaL, particularly in the socio-cognitive-based recognition of the influence of a learner over him- or herself (Bandura, 1991) and the sociocultural acknowledgement of the influence of one's environment over his or her knowledge acquisition and development (Hiebert & Grouws, 2007). Given the apparent clash between the first's orientation around the individual and the second's emphasis on externality, the inclusion of both philosophies in the same framework may seem to some a contradiction. However, TFA blends them by directing the educator to design learning experiences that actively involve each individual learner while taking into account environmental factors (Clark, 2012). The contributions of both sets of principles were captured by Bandura (2006), who declared, "A major goal of formal education is to equip students with the intellectual tools, self-beliefs, and self-regulatory capabilities to educate themselves throughout their lifetime" (p. 10).

Although FA stresses student ownership of their own learning (Falchikov & Boud, 1989), many students do not naturally develop the ability to monitor their learning (Andrade & Valtcheva, 2009). Therefore, another important formative goal is to teach students how to monitor their progress and make plans to improve subsequent work (Clark, 2012; Stiggins, 2009). The teacher is "at the heart" of the connection between assessment and its positive effect on student learning (Stiggins & Chappuis, 2005, p. 12).

In order to truly impact the learner, the educator must know about background experiences, knowledge, and beliefs—in other words, what the learner already "knows—and should adjust assessment and instruction accordingly (Clark, 2012).

Formative-learning goals. The tenets of socio-cognitive and sociocultural theories are highlighted in Pintrich and Zusho's (2007) definition of self-regulation as "an active constructive process whereby learners set goals for their learning and monitor, regulate, and control their cognition, motivation, and behaviour, guided and constrained by their goals and the contextual features of the environment" (p. 741). This definition acknowledges the reality that students' ability to self-regulate is impacted by many factors, such as educational environment, past academic achievement, and comfort with specific content areas. Effective self-monitoring is usually impaired when students dwell on negative personal behaviors or characteristics; are discouraged by the relative difficulty of a task (Kirschenbaum & Karoly, 1977); have a history of low academic achievement (Nicol & Macfarlane-Dick, 2006); are confused by the criteria for successful task completion (Andrade and Du, 2007); or fail to practice effort and perseverance (Butler & Winne, 1995). In addition, students tend to fail to consistently apply self-regulation strategies when they receive feedback that contradicts strongly held beliefs or knowledge, even if the feedback is in fact correct (Butler & Winne, 1995).

Despite factors that can inhibit students' self-regulation, all students can become better-self-regulated learners (Clark, 2012; Nicol & Macfarlane-Dick, 2006; Warner et al., 2012). Much of the responsibility for students' self-regulatory behavior falls upon teachers, whose educational and assessment practices yield beneficial or detrimental effects on students' self-regulation (Ellis & Folley, 2010). Answering this reality, Clark

(2012) decomposed AfL and AaL into eight specific goals to direct teachers as they work to expand students' involvement in the classroom learning and assessment processes and to nurture students' capacity for self-regulation.

Goal 1: Communicating learning goals and criteria for success. Clark (2012) articulated his first goal as to "communicate to students the goals of the lesson and the criteria for success" (p. 222). This goal, prevalent in other research on formative assessment practices (Nicole & Macfarlane-Dick, 2006; Ross, 2006), refers to the need for students and teachers to share comparable ideas of what constitutes quality task completion (Black & Wiliam, 2009; White & Fredrickson, 1998; Wiliam, 2007a). The practice of clarifying formative goals and criteria for success before an assessment is made is sometimes stigmatized as "teaching to the test" (Guskey, 2003); some scholars counter, however, that it enables students to clearly understand the "rules of the assessment game" (Thomas et al., 2011, p. 2) they must play. Andrade and Du (2007) added that "criteria for student work must be so transparent that students can learn to evaluate their own work the same way their teacher does" (p. 160). The result of this transparency will be a classroom with "no surprises, and no excuses" (Stiggins & Chappuis, 2005, p. 3) and students with positive attitudes (Andrade & Du, 2007) and heightened levels of engagement with, interest in, and attention to learning and assessment tasks—students who, consequently, learn more (Ross, 2006).

Before communicating goals to students, teachers must ensure that said goals have qualities representative of a formative focus. In a discussion of goal types, the term *performance* refers to those which involve primarily grades and, by extension, ego and comparisons to others (Black & Wiliam, 2009; Butler & Winne, 1995), such as "getting an A on the test" or "getting the highest grade in the class." Because performance goals are associated with decreased motivation, self-efficacy (Black & Wiliam, 1998a), and academic risk-taking (Heritage, 2010b) and lower self-regulation strategies (Black & Wiliam, 2009), they should be avoided in a formative context.

In contrast to performance goals, *mastery goals*, also called *growth* (Black & Wiliam, 2009) or *learning* (Nicol & Macfarlane-Dick, 2006) *goals*, focus on acquiring and mastering skills and assimilating knowledge (Black & Wiliam, 2009; Butler & Winne, 1995; Clark, 2012; Nicol & Macfarlane-Dick, 2006). These goals are associated with greater student ownership and motivation (Clark, 2012; Perry et al., 2010) and, therefore, should be part of a formative focus.

Finally, so that students are not overwhelmed, formative assessment goals should be relatively short-term, specific, and incrementally work toward broader goals for achievement and self-regulation (Heritage, 2007). Once teachers have aligned their goals with the aforementioned criteria, they have a variety of strategies available to them in the literature by which to communicate goals to students.

Strategies for communicating learning goals and success criteria. One tool widely employed to convey criteria to students is a rubric, which outlines characteristics of various degrees of successful task completion (Andrade and Valtcheva, 2009; Johnson & Svingby, 2007; Moskal & Leydens, 2000). But despite the abundant use of rubrics, evidence exists that even detailed rubrics can be subject to multiple, sometimes conflicting interpretations, potentially leading to miscommunication of the task-completion criteria (Nicol & Macfarlane-Dick, 2006; Rust, Price, & O'Donovan, 2003); other critics have cited that the criteria themselves are too vague or too long and therefore cumbersome (Ross, 2006). Nicol and Macfarlane-Dick (2006) recommended addressing these issues by "providing better definitions of requirements using carefully constructed criteria sheets and performance level definitions" (p. 8), although they did not explain what "better" and more-"carefully constructed" criteria would look like.

Additional tools for conveying goals. Because many rubrics are unclear, unwieldy, or both, educators can employ other techniques to communicate task-completion criteria, including checklists and exemplars. One method some teachers adopt involves issuing students a checklist, which lists requirements in a non-hierarchical way and thus excludes judgments about the importance of any one requirement relative to another (Warner et al., 2012). Another strategy is to increase the amount of class time devoted to discussion of and reflection on criteria, which, previous research has shown, can also raise achievement (Black & Wiliam, 1998a). Other teachers use exemplars and examples of various levels of student work, which has proven especially helpful for those students who can more easily recognize errors in others' work than in their own (Leahy et al., 2005). For that reason, it has also been suggested that teachers facilitate practice sessions in which students provide feedback on each other's work according to predetermined standards (Gielen, Peeters, Dochy, Onghena, & Struyven, 2010).

Andrade and Boulay (2003) tested the effect on achievement of practice sessions where students were guided through the process of applying criteria to their own work. They provided 7th- and 8th-grade students with rubrics that conveyed scoring criteria before having them complete historical fiction essays. The students in the treatment group also received two 40-minute lessons during which they were guided through the process of applying criteria to their own essay drafts. The results showed a significant relationship between the treatment and scores for the girls' essays only. Because the boys' historical fiction essay scores consistently surpassed the girls', the significant effect of practice sessions on girls' scores suggests that the treatment could have a higher impact on the scores of lower-achieving students, a finding consistent with the literature (Black & Wiliam, 1998a).

A study conducted by Kulkarni et al. (2013) tested an intervention's effect on students' ability to accurately apply assessment criteria. The intervention combined the research-based strategies of using a rubric to communicate assessment criteria, providing students with opportunities to practice applying criteria to the work of a peer, and providing students with instructor feedback regarding their application of the criteria. An online class on human-centered design, offered through the University of San Diego, required all of its 5876 students to participate in a type of peer assessment called Calibrated Peer Review. One day after students completed an online assignment, using a provided rubric as guidance, they began the peer-assessment process, which was divided into two parts: calibration and assessment. While engaged in the calibration, students peer-graded an assignment submission, then viewed the faculty grade that had been assigned to the same piece of work the previous day, along with an account from the staff member describing how he or she arrived at the grade; essentially, the students were provided instructor feedback regarding the accuracy of their peer-assessments. If the peer grade was close enough to the staff grade, the student moved on to the assessment part of the process; if not, they repeated the calibration process, up to four more times. At that point, all students moved on to the assessment portion of grading, and students' scores were subsequently compared with instructors'. By the end of the study, which incorporated two cycles of assessments over two assignments, 42.9% of peer grades fell within a 5% difference from instructors' and 65.5% fell within a 10% difference. This study offers evidence that a training series applying recommendations from previous research can more closely align students' assessment skills with those of the teacher.

Including students in developing learning goals. Arguably, the technique that best helps students understand criteria and goals is to include students in the development of said goals and criteria (Andrade & Du, 2007; Heritage, 2007; Nicol & Macfarlane-Dick, 2006) and even in the development of the assessments themselves (Ellis & Folley, 2010; Stiggins, 2009). It has been noted that, rather than simply allowing students to determine on their own what successful task completion should look like, teachers should direct and inform students of the requirements and guidelines that influence curricular considerations of the class (Ross, 2006), such as district and national standards or skills that contribute to future career success.

Student inclusion in the goal- and criteria-development process can be accomplished in a variety of ways. For example, in a classroom activity described in Ross et al. (2002), teachers developed and administered a rich mathematical-problem-solving task to students, who completed the task individually. Afterward, in a whole-class setting, students suggested criteria for judging the quality of their performance and subsequently narrowed their list of suggestions to the four most important characteristics of an excellent response. In small groups, students then described what high, medium, and low performance on the task would look like. Outside of class, the teachers reworked these suggestions and criteria into a rubric using student-friendly language that also reflected the guidelines of the curriculum.

Research has shown that when students are involved in the development of assessment goals and criteria for success, they not only gain a better understanding of standards (Stiggins & Chappuis, 2005) and demonstrate higher levels of subsequent achievement (Andrade, Du, & Wang, 2008), but also receive additional less-foreseeable benefits. These include higher levels of enthusiasm to complete work (Noonan & Duncan, 2005), greater confidence in their work (Stiggins & Chappuis, 2005), elevated engagement due to enhanced buy-in (Andrade & Du, 2007), feelings that they "own" the criteria (Falchikov & Boud, 1989, p. 419), heightened intrinsic motivation (Clark, 2012), and, eventually, the ability to accurately develop high-quality criteria and goals independently (Black & Wiliam, 1998a; Clark, 2012).

Goal 2: Discussing study habits and effective strategies. Once teachers have shared their vision of the learning goals and criteria with students, they must follow by taking steps to help students work toward achieving those goals. One strategy is found in Clark's (2012) suggestion that teachers "engage students in discussions about study habits and strategies which sustain improvement" (p. 222). Many students, especially those who are unmotivated or who have gaps in their knowledge, frequently lack study and organizational skills, such as time management (Clark, 2012) and the ability to select and structure a proper work environment (Credé & Kuncel, 2008). In order to get students to use these skills routinely enough that they begin implementing them without prompting, teachers must do more than describe and model strategies (Nicol & Macfarlane-Dick, 2006); they must help students understand how, when, and why to use them (Butler & Winne, 1995).

Students' prior experiences and beliefs impact their ability to engage with any task, including development of self-regulatory practices (Nicol & Macfarlane-Dick, 2006). Therefore, teachers must use strategies that help them get to know what factors drive their students' behavior. One such strategy, called *interpretive listening*, is to design prompts that elicit students' thinking in a way that teachers may study and better understand students' thought processes (Black & Wiliam, 2009, p. 10). Instructors can also identify and address individual sources of student motivation and needs and then craft tailored study-skills enhancement plans (Kim & Keller, 2008); this individualization, it is thought, may inspire students to find value in improving their learning strategies and, subsequently, to practice and implement them in a more self-regulated fashion (Zimmerman, 2001).

When directly modeling strategies for and reviewing strategies with students, teachers must keep in mind the series of stages that learners typically move through while learning to combine skills with self-regulation (Andrade & Du, 2007; Zimmerman, 2001). The benchmark names used to denote these phases are observation, emulation, self-control, and self-regulation (Andrade & Du, 2007, p. 170). One skill that research has shown to be helpful at all stages in this progression is annotation, "a writing-to-learn strategy for use while reading or rereading. Annotating helps readers reach a deeper level of engagement and promotes active reading" (Porter-O'Donnell, 2004, p. 82), and can be applied to any text in any field, including mathematics. When implemented and practiced correctly, annotation makes learners' thinking visible and promotes higher levels of engagement between learners and texts, as learners identify main ideas, questions, and other connections they might make (Hwang, Chen, Shadiev, & Li, 2011).

As they are with other study skills, students frequently must be taught how to effectively use annotation (Porter-O'Donnell, 2004), as they progress through the aforementioned stages and learn to annotate in a self-regulated way; this is especially the case for students engaged with mathematics texts, because, even when students do not have reading challenges, they frequently struggle with decoding the more symbolic and technical format of many mathematics textbooks (Cantrell, Burns, & Callaway, 2009). While learners are in the earliest stage of acquiring annotation skills, the observation level (Andrade & Du, 2007; Zimmerman, 2001), they watch the teacher or another student model the process and make judgments about the relative quality of the aspects of the model. For example, a teacher might write in the margin of an algebra text a more student-friendly definition of *y intercept*, such as "the place the graph crosses the y axis" or a note illustrating the term through a context familiar to many students, such as an interception in football; the student might then observe that one feature of good annotation is to write something in the margin, rather than just to highlight text. Knowledge of this first stage is important for teachers, and they must be sure to clarify their reasoning for taking particular actions so that students are better able to make and understand these decisions on their own later (Andrade & Du, 2007; Zimmerman, 2001).

In the next stage, students attempt to practice the new technique on their own, applying their interpretation from the model of what constitutes quality. If they are annotating, they might write in the margin, as the teacher model did (Andrade & Du, 2007; Zimmerman, 2001). It should be noted that in order to encourage students to annotate effectively, it is imperative that the teacher monitor and provide feedback to the students as they practice (Butler & Winne,1995), with specific, criteria-referenced suggestions for subsequent improvement (Black & Wiliam, 2009).

As students enter the next, more internalized phase, (Andrade & Du, 2007; Zimmerman, 2001), they will continue to practice annotation while keeping the model in mind; but their implementation will begin to take a shape that reflects what they believe works best for them. For example, they might continue annotating with writing in the margin but adopt a set of symbols that works as a kind of shortcut. Again, teachers should consistently provide feedback (Butler & Winne, 1995). Finally, once they enter the self-regulated phase, learners take on "their own distinctive styles of performing" a skill (Andrade & Du, 2007, p. 171).

Goal 3: Involving students in previewing and planning forthcoming work. In addition to practicing effective study skills, self-regulated students use their background knowledge both to deliberately plan actions they believe will allow them to successfully engage in a given task and to develop conjectures about the progress that will likely be made as tasks progress (Butler & Winne, 1995). By implementing Clark's (2012) suggestion to "involve students in previewing and planning forthcoming work" (p. 222), teachers can help students develop this skill. Previewing content through classroom discussion advances one of the major goals in the CCSSM by shifting students away from the belief that mathematical topics exist in isolation (NGACBP & CCSSO, 2010, Key Shifts in Mathematics section), as students can preview content and speculate about its relationship to previous knowledge. Motivation—and, consequently, learning—is deepened when students are encouraged to make predictions, design plans, and communicate their ideas to others (Blumenfeld et al., 1991). In addition, previewing content with students allows teachers the opportunity to solicit student feedback and gauge their interest in future topics. This outcome, yet another potential source of student motivation, is known as *differentiation*, "an approach to teaching and learning so that students have multiple options for taking in information and making sense of ideas" (Hall, 2002, p. 1).

Goal 4: Seeking help. Self-regulating students strategically seek input from outside sources, including peers (Heritage, 2007; Perry et al., 2010), trusted adults (Ryan, Hicks, & Midgley, 1997), and online curricular materials (Eva & Regehr, 2011), when their expected level of goal attainment does not match what they actually achieve (Butler & Winne, 1995; Clark, 2012). Students with lower self-regulation, however, are neither as aware nor as inclined as higher self-regulators to solicit help when it is needed. In such cases, teachers are directed to "inform students of who can give them help if they need it and permit full access to such help" (Clark, 2012, p. 222). In order to maximize the effectiveness of this strategy, teachers must address the tendency among students with low self-confidence to resist asking for help for fear of looking less intelligent (Black & Wiliam, 1998a; Schunk, 2012), as well as the failure of other students to realize that they need help and the assorted other social motivations that inhibit or prevent students from asking for help (Ryan et al., 1997). To combat this disposition toward help-seeking avoidance, teachers can engage students in discussions about previous experiences where help-seeking resulted in success, identify someone the students trust to be a resource (Wilson & Deane, 2001), or offer students research that demonstrates the positive correlation between academic success and a willingness to ask for help (Clark, 2012). One of the most important strategies, however, is to create a classroom environment built on the concept of collaboration (Heritage, 2007; Perry et al., 2010), which will be elaborated later in the chapter.

Goal 5: Fostering metacognition. Clark's (2012) first four goals, while student-centered, are essentially AfL-oriented and, as such, aim primarily to train the actions of teachers, such as specific discussions they should lead and information they should convey. By contrast, the last four goals are rooted in AaL and look toward developing students as full partners in the classroom. The first of these goals is metacognition.

To fortify students' investment in their learning, Clark (2012) urges teachers to "provide opportunities for students to become meta-cognitive and build knowledge of themselves as learners " (p. 222). As previously mentioned, Clark decomposed metacognition into three major parts: monitoring, planning, and reflecting, which unite to form a "recursive, but not necessarily linear" process (p. 215). While engaging in a task, self-regulating students monitor their progress, judge the degree to which their progress consists with achieving their goals, and make plans for subsequent engagements and goals based on their assessment. Again, feedback—provided by both students and teachers—is at the center of Clark's framework because it is an "inherent catalyst" in the development of all self-regulating strategies, including metacognition (Butler & Winne, 1995, p. 246). Effective feedback should help students answer for themselves (Heritage, 2010a, pp. 79-80) "Where am I going?," "Where am I now?," and "Where to next?," guiding questions that can be applied either to a discrete task or to progress across time (Clark, 2012).

Assessing performance against goals. As noted throughout, students best learn self-regulation when they receive quality feedback from various sources (Nicol & Macfarlane-Dick, 2006). After teachers have conveyed to or collaborated with students to develop learning goals and criteria (Black & Wiliam, 2009; Nicole & Macfarlane-Dick, 2006; Ross, 2006; Wiliam, 2007a), students engage, to varying degrees, in self-monitoring, which is "pivotal to self-regulated learning, the cognitive process that assesses state of progress relative to goals and generates [internal] feedback that can guide further action" (Butler & Winne, 1995, p. 259). Internal feedback, or "information directly perceivable by the learner while task processing" (Narciss, 2008, p. 127), is generated as learners monitor their interaction with a task or process (Butler & Winne, 1995; Nicol & Macfarlane-Dick, 2006) and, therefore, is a necessary product of self-regulation (Clark, 2012). Most self-regulation models agree that monitoring is a conscious process—with very few exceptional cases in which learners' expertise in the content is exceptionally high (Butler & Winne, 1995)—and Sadler (1989) maintained that all students generate internal feedback at various levels. The ability to produce internal feedback and the abundance and quality of the internal feedback produced are dictated by many factors, including learners' sense of self-efficacy and capacity for self-regulation (Clark, 2012, p. 213).

At the same time that students are monitoring their performance, other players, most frequently teachers (Noonan & Duncan, 2005) but increasingly often peers (Sung et al., 2005), also monitor students' performance relative to goals and supply external feedback (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Butler & Winne, 1995; Kluger & DeNisi, 1996). As outlined in previous research, external feedback "is information that helps students trouble-shoot their own performance and self-correct: that is, it helps students take action to reduce the discrepancy between their intentions and the resulting effects" (Nicol & Macfarlane-Dick, 2006, p. 7). External feedback can also be offered in response to students' assessments of their own performance relative to criteria; from a self-regulation viewpoint, external feedback can help students become owners of their own learning by interacting with their internal self-regulation processes and thereby positively impact subsequent learning outcomes (Andrade & Valtcheva, 2009)

Consistent with the sociocultural tenet of scaffolding for students within their zone of proximal development (Heritage & Heritage, 2013) is the notion that students' learning outcomes are improved by this internal and external feedback interaction. Feedback interactions have been shown to help students calibrate the accuracy of their self-monitoring process and make adjustments if needed (Butler & Winne, 1995; Lawson et al., 2012; Nicol & Macfarlane-Dick, 2006). In order for this calibration to occur, students must engage with the feedback in a conscious, meaningful way, as they negotiate feedback with current beliefs and understandings to construct new knowledge (Bangert-Drowns et al., 1991; Butler & Winne, 1995; Nicol & Macfarlane-Dick, 2006).

The literature demonstrates that all students, even the very young, can monitor effectively (Butler & Winne, 1995; Perry et al., 2010), but gaps between students' and teachers' conceptions of learning goals and successful-task-completion criteria (Andrade & Du, 2007; Nicol & Macfarlane-Dick, 2006) can sometimes interfere with this process. Said gaps may stem from a variety of factors, including too much criteria (Butler & Winne, 1995), the failure to develop a teacher-student aligned interpretation of criteria, (Heritage, 2007), a misunderstanding of the task (Butler & Winne, 1995), learner failure to value the task (Black & Wiliam, 2009), and learner overconfidence (Butler & Winne, 1995).

Students' verbal or written articulation of their internal feedback on their progress (Clark, 2012) can help teachers identify the source of inconsistent conceptions of tasks (Nicol & Macfarlane-Dick, 2006, p. 9). To draw students' explanations, teachers can prompt students to justify their reasoning for a particular solution or belief (Stepanek & Jarrett, 1997), encourage student reflection, or offer students topics to discuss with classmates (Wiliam, 2007a). The power of this dialogue was enunciated by Voogt and Kasurien (2005): "Formative assessment may consist of hard data, but more often and more importantly of 'tacit knowledge', i.e. knowledge that both the teacher and student obtain through discussion, reflection and experience" (p. 8). In the context of fostering self-regulation, feedback will generally serve one or more of the following purposes: to confirm understandings or beliefs; to fill incomplete understandings of information or gaps in knowledge; to replace misconceptions with correct information; to sharpen or clarify information that is mostly correct; or, if a student has deeply rooted misconceptions, to break down and rebuild all previously held knowledge of a particular area or field (Butler & Winne, 1995)

Characteristics of quality external feedback. From a self-regulatory view, effective feedback is a "critical feature in determining the quality of learning activity and, therefore, a central feature of pedagogy" (Black & Wiliam, 2009, p. 2). And quality external feedback, as noted above, participates in the feedback cycle that fuels student self-monitoring. When considering the characteristics that maximize the effectiveness of external feedback, one must address the feedback's content, delivery, and the aspects of student performance at which it is directed.

Feedback can be delivered via remote methods, such as by computer programs (Nicol & Macfarlane-Dick, 2006) or text (Porter-O'Donnell, 2004), and still contribute to student self-regulation. *Cognitive validity feedback* is received via the connection a learner makes between a task-prompt characteristic and an implication for successful task completion. For example, if a mathematics-test study guide at the end of each problem requires that students explain how they obtained their answer, they may conclude that they will receive a better grade on the test if they recount their thought process (Butler & Winne, 1995). Scholars have found that this kind of feedback helps the learner distinguish those pieces of information most important to optimally accomplishing a task.

Research suggests that when external feedback contains a score or a grade, students tend to devote time and energy to comparing themselves and their work with that of other students, rather than to developing ways to revise and make improvements to it (Nicol & Macfarlane-Dick, 2006; Wiliam, 1999). Even a score accompanied by suggestions for improvement distracts students from addressing how they might improve their work (Wiliam, 2007a). Therefore, if external feedback is to serve a formative purpose, it should not include a grade but comments only (Black & Wiliam, 2009).

Effective external feedback cannot simply deliver praise (Kluger & DeNisi, 1996) or nonspecific judgment or guidance, such as "try again," that fails to make clear to students how or why they should improve work or continue along the same successful path (Bangert-Drowns et al., 1991; Clark, 2012). An example in this vein is *outcome feedback*, which describes performance or information only as correct or incorrect.

Though common, outcome feedback does very little to promote self-regulation (Butler & Winne, 1995), has been shown to negatively impact achievement (Bangert-Drowns et al., 1991), and, according to most scholars, should be avoided. To be effective, external feedback must be directly related to predetermined criteria and provide specific evidence for evaluation (Nicol & Macfarlane-Dick, 2006), for staying the same successful course (Black & Wiliam, 2009), and/or for subsequent improvement, tailored to the specific kind of error a student is making (Butler & Winne, 1995; Clark, 2012; Leahy et al., 2005; Nicol & Macfarlane-Dick, 2006).

Quality external feedback should always be directed toward work and performance, never toward imputed characteristics of the students themselves (Gielen et al., 2010). Feedback pitched at personal characteristics of the students, rather than at their work, has been shown to impede subsequent performance, even when the feedback is positive (Kluger & DeNisi, 1996; Wiliam, 1999) Similarly, effective external feedback also involves no comparisons with peers (Black & Wiliam, 1998b) and instead is always, again, criteria-referenced.

It is not necessary that feedback that heightens students' ability to revise their products be confined to explicit suggestions for improvement. *Functional validity feedback* compares learners' assessment of success in task completion with their actual success. A student might convey during a reflective writing exercise she felt she knew the answers to less than half of the problems on a recent quiz, and she receives functional validity feedback when the teacher tells her that she actually answered 70% of the questions correctly (Butler & Winne, 1995).

Even high-quality, ideally constructed external feedback can lose its effectiveness if students become overwhelmed by what they perceive to be too much information dispensed at once (Butler & Winne, 1995). Therefore, teachers should limit feedback in amount and scope (Nicol & Macfarlane-Dick, 2006; Wiliam, 1999). In situations that require an abundance of wide-ranging feedback, the areas of highest priority should be addressed first (Nicol & Macfarlane-Dick, 2006).

Planning. Once students have monitored their progress and (in many cases) received commentary on their self-generated feedback, they enter the planning stage, marked by account-taking of what they do or do not know (or what they do or do not know how to do) and plan-forming for subsequent work or task engagement (Clark, 2012; Nicol & Macfarlane-Dick 2006). As this plan unfolds, they also generate internal feedback, which can induce a change in any number of students' internal phenomena, including interpretation of the task, motivational beliefs, or domain knowledge, any or all of which may impact the student's self-regulation (Nicol & Macfarlane-Dick, 2006). The most common use of this internal feedback is the development of a plan for subsequent learning progressions over the same topic (Clark, 2012). And just as teachers' feedback should be specific, short-term, and limited in scope to prevent students from feeling overwhelmed, so should students' goals and plan have the same qualities (Heritage, 2007). The more that students actively engage in this future-planning process, the more self-regulated they become (Andrade & Du, 2007).

A vital part of students' planning process is teacher feedback, which should strive to help students calibrate the appropriateness of goals and develop ones that allow them to move forward. Many students, owing to previous experiences, beliefs, and knowledge, will have difficulty formulating appropriate goals and the strategies by which to accomplish them (Clark, 2012); teachers can help students craft more-effective plans for subsequent improvement by modeling strategies (Nicol & Macfarlane-Dick, 2006). As a cautionary note, if students see someone modeling a strategy and are unable to relate to that model, they may misapprehend that the strategy does not apply to them (Zimmerman, 2001). Therefore, teachers should share with students their own struggles as well; McGregor (2007) has developed classroom activities that allow teachers to model these for students.

Self-regulated students are aware of their knowledge base; in other words, they know what they know (Butler & Winne, 1995). When faced with external feedback that contradicts previously held knowledge or beliefs, ideally, students generate feedback that guides them toward action to internalize and apply the new assessment. However, research shows that this is not common (Butler & Winne, 1995). Given this, feedback should focus more on learning goals. This plan can also be affected by the learners' view, after monitoring, of the amount of effort that subsequent plans may entail (Butler & Winne, 1995, p. 259).

Reflection. Students may reflect on their progress over an extended period of time or with a particular topic or set of criteria (Andrade & Du, 2007, p. 160). Reflection frequently comes after either monitoring or planning (Clark, 2012). Reflection also occurs after students have compared their actual progress with their anticipated progress; if what was earmarked as adequate progress has not been achieved, better-self-regulating students either revise the original goal, the plan to achieve the goal, or the set of previous knowledge and beliefs that informed the original development of the goal (Butler & Winne, 1995).

In sum, reflection demands that students review their history of strategies that did and did not result in task-completion success. By this, students are able to build and record a larger set of strategies from which to draw when they next have a related task to undertake (Black & Wiliam, 2009). Thus equipped with this menu, students can better adapt to, or formulate new strategies appropriate for, new learning situations as they present themselves (Butler & Winne, 1995).

Reflection—and thus self-monitoring—may be aided by active record-keeping of thought processes. Thus, one way for teachers to facilitate reflection is to encourage and model various record-keeping strategies (Clark, 2012) whereby students can document and track and their progress and reflection over time (Zimmerman, 2001). Journals, in which students regularly mark and reflect on their progress, have been shown to increase autonomy and self-regulated learning (Clark, 2012).

Goals 6, 7, and 8: Establishing a learning environment. Stated plainly, Clark's (2012) final three goals unite under the single banner of creating a classroom environment that nurtures self-regulation. The goals as Clark articulated them involve "[c]reat[ing] a non-comparative, productive environment free of risks to self-esteem founded upon cooperation and dialogue," "support[ing] students as they take more responsibility for their learning," and "provid[ing] opportunities for frequent participation in the process of learning, with the teacher as their advisor and with their peers in a climate of equality and mutuality" (p. 223). Classrooms that foster this kind of environment feature students who, rather than immediately run to the teacher for assistance, more readily search their own repertoire of strategies or seek out, strategically, peer assistance (Perry et al., 2010) when they are faced with a new, challenging situation. In short, these goals foster an environment where all participants are valued and the improvement of students' sense of self-efficacy is paramount (Butler & Winne, 1995, p. 256).

Benefits of self-efficacy. Clark's (2012) emphasis on self-efficacy emerges from the well-established body of literature that connects self-efficacy, self-regulatory behaviors, and achievement (Nicol & Macfarlane-Dick, 2006). Students' self-efficacy directly influences the kind of goals they set, their motivation to complete goals, their decisions about what to do when obstacles are encountered (Butler & Winne, 1995), and their level of productivity; it also increases the likelihood that they will work to apply corrective external feedback (Clark, 2012). Those with heightened self-efficacy are more likely to be optimistic in their predictions about task completion, to set high expectations for themselves, to embody greater levels of perseverance (Ross, 2006), and to view mistakes as a non-threatening part of the learning process (Perry et al., 2010).

Effects of existing characteristics. When students first enter the classroom environment, they bring with them preexisting levels of self-efficacy that have been shaped by multiple factors, both academic—such as comfort with and knowledge of a particular subject or content area—and personal—such as attitudes toward their teachers, school, and education, perseverance, and responsibility (Heritage, 2007). As Stiggins noted (2009), students frequently use their assessment experiences and the perceived results to inform their view of themselves as students, which can in turn affect self-efficacy. Students' views can impede their ability to compare their actual level of performance with criteria; those more prone to focusing on their failures automatically assess themselves at a lower rate (Kirschenbaum & Karoly, 1977), while, at the opposite end of the spectrum, overconfidence can also lead to poor self-regulation (Butler & Winne, 1995). According to Black and Wiliam (1998a), learners' views of their own abilities can act as a self-fulfilling prophecy. Ideally, therefore, students will learn to "approach difficult tasks as challenges to be mastered, rather than as threats to be avoided" (Stiggins, 2009, p. 420).

Developing collaborative characteristics. When teachers use collaborative activities and student-centered discussions that provide opportunities for quick feedback and students' active engagement, self-efficacy is elevated, because students see the importance of their participation in their own learning and their voice in the construction of knowledge (Clark, 2012; Nicol & Macfarlane-Dick, 2006). One specific method of building a collaborative environment may be borrowed from a study of Japanese group discussions. These are structured to value all students' opinions with wording that is specifically used to nurture the idea of a "collective student" (Black & Wiliam, 2009, 24); the implication is that individual contributions are reflective and determinative of the entire class's learning. Previous literature also suggests that, when people engage in cooperation, portions of the brain associated with reward processing are stimulated, which further promotes the notion that a collaborative learning environment contributes to motivation and self-efficacy (Rilling et al., 2002).

In a collaborative classroom, there exists a balance between students' voices and the teacher's voice, which provides necessary direction and structure (Black & Wiliam, 2009). In order to set up a safe environment in which students feel comfortable taking academic risks, the teacher must establish norms. As part of that process, students must be explicitly taught and modeled how to express their thinking, to listen to each other, to disagree in a respectful manner, and to provide constructive feedback (Heritage, 2007). Many students must be trained in the areas of productive collaborative group work, which they practice as they plan a group-oriented task and negotiate through any disagreements (Clark, 2012).

The effort expended to teach students how to engage productively in dialogue is rewarded, given both the aforementioned benefits to students' self-efficacy and the higher achievement that students often show (Butler & Winne, 1995). Because the teachers enjoy an enhanced understanding of students' thinking and knowledge (Clark, 2012), they are able to more effectively adjust instruction when appropriate (Nicol & Macfarlane-Dick, 2006).

Creating shared responsibility. When a classroom environment is collaborative and students share the responsibility for learning, an implicit sense of trust develops between the teacher and students (Stiggins & Chappuis, 2005). Increases in SRL have been seen in classrooms where teachers value students' input into decisions such as the format of learning activities, the level of difficulty of activities, and the method they use to demonstrate understanding (Perry et al., 2010). Students experience ownership of learning experiences and feel as if, regardless of skill level, they can contribute and be valuable sources of learning for one another (Heritage, 2007). As a result, they develop a greater sense of commitment to classroom learning activities (Stiggins & Chappuis, 2005), a commitment that manifests itself in heightened levels of engagement and perseverance and that, in turn, boosts capacity for self-regulation (Clark, 2012).

Mastery goals. Self-efficacy and subsequent learning gains blossom in classrooms that focus on aforementioned mastery goals, rather than on performance goals (Butler & Winne, 1995; (Noonan & Duncan, 2005; Perry et al., 2010), because they engage in learning and not mere evaluation (Black & Wiliam, 1998a). Aspects of the environment and task, and unique personal factors influence students' decision to approach particular goals from a mastery or ego/performance view (Zimmerman, 2001), but this decision can be shaped. An ideal environment allows students to revise and resubmit work (Guskey, 2003; Nicol & Macfarlane-Dick 2006; Ross, 2006; Stiggins, 2002); if resubmission after grading is not a possibility, students should be given multiple opportunities to receive feedback during a formative cycle, before final work completion or submission (Andrade & Du, 2007; Nicol & Macfarlane-Dick 2006). Policies for revision and resubmission, Clark (2012) argued, will induce students to buy into the notions that learning happens incrementally over time and that they often will not demonstrate mastery the first time a task is tackled. Without opportunities to resubmit work, it is likely that students will not see the value in improving it (Zimmerman, 2001). It serves an assessment purpose for teachers as well, according to Boud (2000):

> The only way to tell if learning results from feedback is for students to make some kind of response to complete the feedback loop. This is one of the most-often-forgotten aspects of formative assessment. Unless students are able to use the feedback to produce improved work, through for example redoing the same assignment, neither they nor those giving the feedback will know that it has been effective. (p. 158)

This focus on mastery over performance and on self-efficacy should also be seen in feedback that is not made public, values mistakes, and offers specific ways to improve (Black & Wiliam, 1998a).

Effort and persistence. Students' levels of effort, persistence, motivation, and ambition derive primarily from their predispositions and past experiences. In order for students to develop the belief that success is possible, they must have opportunities to experience real success with a task considered challenging; this can develop into greater confidence over time (Stiggins & Chappuis, 2005). For those students who have yet to master the challenges appointed for them, the occasion to witness models of overcoming adversity can be a step toward their future mastery (Zimmerman, 2001).

Self-regulating students are able to effectively weigh the potential costs of working toward goals against the gains that would accrue from realizing the goals. In other words, they pair their expectations of the amount of effort successful task-completion requires and their actual progress with the task (Butler & Winne, 1995) and gauge their perseverance accordingly. If a task is particularly adverse, students may choose to avoid similar tasks in the future, even if there was eventually success, because, for them, the success was exceeded by the expected effort (Butler & Winne, 1995). Students can actively work to increase their motivation through strategies such as self-consequences and self-verbalization (Clark, 2012).

The literature identifies a phenomenon called a "just right" gap (Heritage, 2007, p. 145), which refers to the size of the gap between a goal and students' level of performance. If students perceive the gap as too great, they likely will feel that attempting to close it is hopeless (Zimmerman, 2001); if they perceive it as too small, they might find no purpose in attempting to close it (Heritage, 2007). Self-efficacy normally determines whether the student will attempt to overcome the obstacle (Butler & Winne, 1995), but even confident students will sometimes, upon evaluating their performance, suffer a blow to their self-efficacy (Zimmerman, 2001). To address these problems, teachers must explicitly convey the value of every task (Heritage, 2007) and teach students volitional strategies, "metacognitive knowledge to interpret strategy failure and knowledge of how to buckle down to work" (Clark, 2012, p. 214), which are positively correlated with persistence (Black & Wiliam, 2009).

Self-Assessment

Tied to students' ability to self-regulate is their ability to engage in self-assessment, when they "make judgments about their own achievement and learning processes, and decisions about action they need to take to make further progress in learning. In order to do this, they need to have a clear grasp of the goals of the learning and of the criteria to be applied in judging how well the goals have been attained" (Deakin-Crick, Sebba, Harlen, Yu, & Lawson, 2005, p. 3). Self-assessment occurs as students evaluate their own perceived current level of performance against pre-established criteria for success (Andrade & Du, 2007; Nicol & Macfarlane-Dick, 2006; White & Fredrickson, 1998). Black and Wiliam (1998b) enumerated the three stages of this process: "recognition of the desired goal, evidence about present position, and some understanding of a way to close the gap between the two" (p. 85). The use of self-assessment, described by Clark (2012), to train students in this process and make them more active participants in their own learning has been increasingly adopted (Sung et al., 2005).

Self-assessment has been demonstrably beneficial to students of various ages and across subject content. Indeed, benefits have been seen with students in elementary, middle, (Ross et al., 2002), and high school (Noonan & Duncan, 2005), and higher education (Falchikov & Boud, 1989), across all major core content areas (Falchikov & Boud, 1989; Noonan & Duncan, 2005), social and physical sciences (Falchikov & Boud, 1989) and in developing students' positive group interactions (Ross et al., 2002).

Self-assessment, self-regulation, and achievement. Self-assessment has been called a "prequel to self-directed learning activities that will improve overall performance and thereby maintain competence" (Eva & Regehr, 2011, p. 312), because of its link to increases in self-regulation (Andrade & Du, 2007; Eva & Regehr, 2011; Heritage, 2007; Nicol & Macfarlane-Dick, 2006). As noted previously, students who take ownership of their learning by setting goals and criteria for success and monitor their own progress toward reaching them—in other words, those who self-assess—are generally higher-achieving than those who do not (Andrade & Valtcheva, 2009; Black & Wiliam, 1998a; Brown & Hirschfeld, 2007; Nicol & Macfarlane-Dick, 2006).

Self-assessment has been shown to promote academic achievement in many contexts within elementary, middle, and high school (Black & Wiliam, 1998a) and in higher education (Nicol & Macfarlane-Dick, 2006) and across content areas, including mathematics pre-to post-tests, reading rates, physics, algebra (Black & Wiliam, 1998a), writing of historical fiction essays (Andrade & Boulay, 2003), college final exams (Nicol & Macfarlane-Dick, 2006), and standardized test scores (Noonan & Duncan, 2005). It follows, then, that students' ability to make accurate judgments about their own work is an indispensable part of their becoming "lifelong learners" (Falchikov & Boud, 1989, p. 395) and "accomplished and effective professional[s]" (Thomas et al., 2011, p. 2).

Student attitude. The correlation between self-assessment and achievement likely proceeds from the association between self-assessment and other qualities that underlie achievement. Students who are encouraged to participate in self-assessment tend to have more positive attitudes toward assessment (Black & Wiliam, 1998a) and display greater levels of engagement and buy-in to assessments, largely because they feel more included in the assessment process (Ross, 2006). They also exhibit greater motivation (Andrade & Du, 2007; Noonan & Duncan, 2005) and persistence (Black & Wiliam, 1998a; Noonan & Duncan, 2005) and superior critical thinking skills, (Andrade & Du, 2007).

Foundation in formative learning theory. Black and Wiliam (1998b), asserted that "self-assessment by pupils, far from being a luxury, is in fact an essential component of formative assessment" (p. 85). Several aspects of self-assessment explain why it is widely viewed as a formative learning experience and, in turn, a source of student self-efficacy (Noonan & Duncan, 2005).

As with other formative-assessment practices, in order to maintain the integrity of the self-assessment process, students should not be responsible for assigning themselves a grade (Boud & Falchikov, 1989). Without grades, students are encouraged to treat assessment as an opportunity for improvement and as a learning experience (Ross, 2006) and become, over time, more willing and able to find and address their own mistakes (Noonan & Duncan, 2005) and to record and track their growth (Stiggins & Chappuis, 2005). When coupled with teachers' goals that promote student involvement in their development, certain qualities of self-assessment—specificity, attainability, a focus on short-term actions with long-term implications, documentation, and a focus on a growth—make students more likely to expect success during future assessments (Ross, 2006).

Role of external feedback. Typically, the audience for student-self-assessment feedback comprises both the students themselves and their teachers (Sadler, 1989). This is because student self-assessment often involves negotiation between internal feedback and teacher- or peer-provided external feedback (Narciss, 2008; Nicol & Macfarlane-Dick, 2006; Stiggins & Chappuis, 2005). As stated above, the latter form of feedback is part of the project to build students' ability to effectively self-assess and often takes the form of comments or questions crafted for students to guide their thinking as they self-monitor (Clark, 2012).

It has been argued that students, with varying degrees of success, already make self-assessments unprompted by teachers (Sadler, 1989). Therefore, teachers should maximize the effectiveness of these assessments by providing structured opportunities for students to self-monitor, assess the degree to which they have met their goal, and reflect on their progress (Nicol & Macfarlane-Dick, 2006). External feedback does not have to be verbal or even emerge directly from a teacher or peer for it to interact with students' internal feedback. For example, students may be required, as they engage in self-assessment, to use colored pencils or highlighters to underline in a rubric a characteristic of successful task completion, then use the same color to flag examples of the applicable characteristic in their own work. They may then make written notes of the rubric characteristics that are missing from their work (Andrade & Valtcheva, 2009).

Effective enhancement of self-assessment has been linked to increased levels of achievement. For example, in Sadler and Good's study (2006), students self-assessed a science test, according to pre-established criteria, received written feedback from the teacher regarding the accuracy of their assessment, and were, subsequently re-administered the test, with no advance notice. Students who self-assessed improved dramatically on the retest, significantly more so than students who had engaged in peer assessment of the same test or no assessment at all. It was speculated, based on these results, that self-assessment greatly enhances student learning.

Self-assessment tools. When employed for self-assessment, writing activities can be valuable sources of information for both teachers and students. When students are directed to describe or reflect on a task, they frequently must process their thinking more than if they were simply performing a procedure. And when teachers supervise this processing, they can gain a much deeper understanding of students' thinking. As one example, students may analyze their answers on a math test and provide a written justification for why their answers are correct or require modification (Andrade & Valtcheva, 2009; Stepanek & Jarrett, 1997); as another, students may use a journal to reflect on their understanding of a topic featured in a lesson or record their thinking while solving a mathematical task (Suurtamm et al., 2010). Students may also pinpoint, before receiving any outside feedback, what they believe to be their strengths and possible opportunities for improvement and explicitly ask a reviewer for specific kinds of outside feedback (Nicol & Macfarlane-Dick, 2006). In other classrooms, students may choose from a list of possible activities that are differentiated according to students' own self-selected level of understanding; students who feel they do not understand choose the red activity, while students who feel they do understand the activity choose the green, etc.) (Suurtamm et al., 2010).

A common tool for guiding students through self-assessment is the rubric. But, as noted, even if students are evaluating their work against a rubric, the format of the rubric may affect the purpose and result of the self-assessment. As in the example cited previously, students may review their answers on a math test and submit a written explanation for why their answers are correct or entail revision (Andrade & Valtcheva, 2009; Stepanek & Jarrett, 1997) before the work is actually turned in (Clark, 2012). This relationship between self-assessment purpose and rubric format may be most especially apparent when students must resort to a more generic rubric, like that issued during state-mandated standardized testing. An example of this kind of generic rubric can be found when Kentucky students take the Kentucky Performance Rating for Educational Progress (K-Prep) assessment. At the start of the test, students are given a generic rubric by which to assess their work, and students may believe, due to ambiguous wording (Ross, 2006), that they have met the second criterion—"You demonstrate in-depth understanding of the relevant concepts and/or processes" (Kentucky Department of Education [KDE], 2014c)— even when their answer does not in fact meet this criterion. In a contrasting case, a student may believe that she has labeled the axis of a graph correctly, until she sees the axis labeled on a rubric much more specifically developed to ascertain successful performance of one particular mathematical task. Here, the student profits from the specificity of the rubric and understands how her performance aligns with the demands of the task.

Peer Assessment

As educators have increasingly recognized the value of social interaction in learning, peer feedback and assessment have become more prevalent (Sung et al., 2005). Opinions vary as to whether peer assessment is a standalone strategy or one that is meant specifically to complement self-assessment (Noonan & Duncan, 2005), but it is commonly agreed that peers can be an assessment resource when they "provide either feedback or grades (or both) to their peers on a product, process, or performance, based on the criteria of excellence for that product or event" (Thomas et al., 2011, p. 3). Peers can also contribute to developing assessment criteria and reflection, and help other students convey, verbally and in writing, what they have learned (Sung et al., 2005). Peer assessment is thought to be most effective when integrated early in a learning process (Noonan & Duncan, 2005) and, as with other forms of formative assessment, when students are permitted to use the assessment feedback to make revisions to their work (Nicol & Macfarlane-Dick, 2006). Research shows that peer assessment, like self-assessment, has a positive impact on achievement, across a variety of contexts and content areas (Clark, 2012; Nicol & Macfarlane-Dick, 2006).

Peer assessment and self- and other-engagement. The literature has documented several formative gains that students make through peer assessment. Among these is that students, sometimes blind to opportunities for improvement in their own work, often more easily identify such instances in the work of others (Leahy et al., 2005, p. 8). By evaluating their peers' work, then, students acquire skills in objectively applying criteria, which they can later use when reviewing their own work (Nicol & Macfarlane-Dick, 2006). Moreover, the feedback providers are able to practice and therefore sharpen their evaluation skills without the negative emotions that can sometimes come while analyzing one's own product (Wiliam, 2007a). It has also been noted that when students receive evaluative feedback from a peer rather than from an instructor "expert," they are more likely to critically analyze and engage in discourse about the feedback (Gielen et al., 2010). Finally, some students more readily accept critiques from fellow students than from instructors (Nicol & Macfarlane-Dick, 2006).

A second gain is that, once students become accustomed to peer feedback, they acquire a greater appreciation for it (Wagner et al., 2011) and request more (Onion & Javaheri, 2011). This appreciation, in turn, feeds student buy-in, a greater focus on mastery goals rather than performance (Noonan & Duncan, 2005), and enhanced levels of perseverance (Nicol & Macfarlane-Dick, 2006) and self-regulation (Sung et al., 2005)—all of which, as previously established, promote student learning. This may be complemented by a teacher's willingness to share the assessment process with students, which has been shown to activate and increase students' enthusiasm, buy-in (Noonan & Duncan, 2005), motivation (Nicol & Macfarlane-Dick, 2006), and sense of self-efficacy (Ross, 2006).

Learning processes. Peer assessment likewise plays a significant role in the training of students' reasoning processes and intellectual expansion. Nicol and Macfarlane-Dick (2006) observed that students construct new learning when they negotiate their previous knowledge and beliefs with contradictory ones of their peers (p. 9). The signal characteristic of peer feedback, the absence of an "expert," frequently leads to elevated discourse that would not occur with teacher feedback alone (Gielen et al., 2010). The importance of this process is highlighted by the development of Common

Core State Standard for Mathematical Practice 3, to "[c]onstruct viable arguments and critique the reasoning of others" (NGACBP & CCSSO, 2010).

Self- and Peer Assessment in the Classroom

Controversy among teachers and students. Considering the various, well-documented benefits of self- and peer assessment, it should follow that teachers happily introduce these strategies in their classrooms. However, studies have shown mixed feelings among teachers and students alike about its use. Many teachers, for example, are hesitant, because of contradictory evidence that raises doubts about students' ability to accurately self-assess (Eva & Regehr, 2011), and interviews have revealed especially large doubts that lower-achieving students can assess higher-achieving students (Bryant & Carless, 2010). This mistrust of student evaluation is not limited to teachers; some students themselves fear dishonest grading (Ross, 2006) and have also indicated mistrust in their peers to take the assessment process seriously (Bryant & Carless, 2010). MacDonald's 2011 analysis of open questionnaires from 52 nursing students who had engaged in self- and peer-assessment activities during their recent modules uncovered mixed feelings among students toward peer-assessment practices. On the positive side, students indicated beliefs that their peers could be sources of learning, that peer assessment offered motivation to perform well, and that students could fairly and truthfully assess themselves and others. However, they also worried about the presence of bias in the process, a lack of confidentiality, and high levels of pressure due to feelings of responsibility that could be caused by peer assessment.

Most researchers equate self-assessment accuracy with teacher assessment agreement (Falchikov & Boud, 1989). Given the fact that there is sometimes disagreement in scoring even among experienced teachers, this correlation may be problematic, but it is the one most consistently used (Falchikov & Boud, 1989). It has been argued that, because of the gains associated with self-assessment, energy devoted to examining the inconsistencies between students' and teachers' grades would be better spent building students' capacity in formative assessment (Boud & Falchikov, 1989). However, it is worth investigating ways to improve student accuracy in order to address the aforementioned fears among teachers and students and to enable them to reap the benefits of more-regular use of peer and self-assessment.

In order to improve the perceived accuracy of self- and peer-assessment strategies and, thus, the likelihood that they will be adopted, it is necessary to examine the research that concerns teachers and students to pinpoint trends therein. Boud and Falchikov (1989) conducted a review of 48 quantitative studies that examined the discrepancy between teachers' and students' assessments. Although they perceived some trends, such as more-accurate self-assessment by students who had been in school longer or who were higher-achieving, overall, they found it difficult to draw many conclusions. They attributed this to the lack of clear assessment scales or consistent sets of criteria used by both teacher and students within the same study or across studies and to the inclusion of subjective criteria, such as effort; similarly, they questioned whether teachers' assessments should have been the benchmark by which accuracy is judged, especially as students' levels of thinking increased. The researchers argued that many of the studies did not truly represent self-assessment, because students were not involved in the process of developing the criteria but merely applied criteria that were given to them; therefore, the authors contended, students were only self-marking, not self-evaluating. Neither did the studies indicate whether repeated practice of self-assessment honed students' assessment skills over time or whether there were gender-based differences in said skills. And the trends that the authors were able to abstract, in the more limited category of self-marking, might be considered suspect by some. Still, two observations are worth noting: first, higher-performing students and students in more-advanced classes were more realistic about, or even underestimated, themselves in their assessments; and second, students as a whole overestimated their grades.

Falchikov and Boud (1989) conducted a follow-up meta-analysis of 57 studies and, in many ways, confirmed the few possible findings of their previous analysis. For example, they also found that when effort level is included as an assessment standard, the gap between teachers' and students' respective assessments widens; that students in higher-level classes tended to have assessments that were more consistent with their teachers'; and that less-subjective subjects, such as mathematics, tended to yield more consistent assessments. And as with the previous study, their authors expressed concerns about unclear assessment methods and criteria. They did observe, however, that when assessing within non-social sciences and when criteria was shared more explicitly with students, student self-assessment was more consistent with the teachers.

Subsequent studies have supported some of Falchikov and Boud's conclusions and cited additional ones. Andrade and Du (2007) backed the earlier observation that when assessments asked students to predict their grades, students tended toward inflation. Wagner et al. (2011) found no difference between pharmaceutical students' self- and teacher assessments of course grades, they speculated, due to clearly conveyed assessment criteria. Ross (2006) found that students, and especially younger students, often failed to understand what quality work looks like or how to apply predetermined standards to their work. Fox and Dinur (1988) saw greater agreement between self-assessment and teacher assessment when students were made aware that their assessments would specifically be compared with teacher or peer assessments. Narciss (2008) revealed that for higher-level cognitive tasks, students are less accurate in predicting whether their answers are correct. Pakaslahti and Keltikangas-Järvinen (2000) uncovered that students are more accurate in their assessments when the criteria are more objective, such as correct/incorrect.

Topping (1998) analyzed 109 previous peer assessment studies, 31 of which involved comparing teacher and peer assessments. Of the studies, 72% reported acceptable levels of reliability, although there was some inconsistency about whether peer or self-assessments were more reliable; there was also a great deal of variance in reliability according to assessment type. It was noted that although more studies reported high reliability of peer assessment, it could be that the less-reliable studies were simply not reported, casting doubt on any claim that peer assessment is in fact more reliable. The review, like the analyses of self-assessment, also reported that consistency of study design and classification of constructs were obstacles to drawing reliable conclusions in some cases.

Aligning expectations and results. Despite the mixed degrees of reported agreement between student and teacher assessments, research has shown that certain self- and peer-assessment training sequences and enhanced collaboration among scorers can produce student scores that are more consistent with teacher assessments. One example is the aforementioned study of students conducting peer assessment through an online class

(Kulkarni et al., 2013). In another example, Sadler and Good (2006) compared teacher grades with self- and peer-assigned marks of 386 tests in four 7th-grade science classes. Before they began grading, students developed, as a whole group, a rubric that would be used to assess the tests, and the teacher discussed with each class criteria that had emerged from other classes' discussions. During the actual self- and peer-assessment process, students were encouraged to discuss their assessments as they compared tests with the rubric and sought feedback from peers. When the researchers examined the relationship between peer- and self-assessment and teacher-generated grades, they found correlations of 0.905 and 0.976, respectively.

Lawson et al. (2012) involved 239 undergraduate business students in an examination of an intervention designed to enhance accuracy of self-assessment of desirable graduate characteristics, including business knowledge communication, attitudes and values, and critical thinking. Students first were provided with the criteria by which they should assess these desirable traits, then were asked to compare their performance against assessment criteria and assign a mark. Instructors then assessed students and could see the self-assessment, which provided them a source of feedback to offer students. Students were then able to view the instructor's assessment and feedback in order to compare their own assessment with the instructor's. In keeping with the recommendations of previous research, students were given feedback before they were told the instructor's score. The initial self-assessment showed a large gap between students and teacher evaluations, with students overestimating their performance. When students engaged in a second round of self-assessment, however, there was no significant difference between their scores and those of instructors. By the end of the sequence, students conveyed a better awareness of desirable graduate attributes as well.

Ross, Hogaboam-Gray, and Rolheiser, (1999) analyzed the results of self-assessment training on self-assessment accuracy of 4th-, 5th-, and 6th-grade students' narrative writing over an 8-week period. First, students engaged in four 30-minute lessons during which they collaborated with the teachers to develop assessment criteria, which the teachers used to create a rubric. The teachers subsequently offered illustrations of various levels of narrative-writing quality to model for students the process of applying criteria; the students were then allowed time to practice applying the criteria themselves. Teachers also conferenced with students about the accuracy of their assessment evaluations and encouraged them to set goals for their future writing. Finally, throughout the training period, students engaged in twelve 3-5 minute practice sessions during which they evaluated their own narrative writing. At the end of the training period, treatment students more accurately evaluated their own work than control students.

Influence of the Literature on the Study Intervention

Feedback: The Center of FA

At the center of Clark's (2012) framework is feedback, which he called "pivotal" to FA strategies (p. 207). Within an educational setting, where learning sequences are based on goals and standards, feedback is "information about how the students' present state (of learning and performance) relates to these [specific] goals and standards" (Nicol & Macfarlane-Dick, 2006, p. 1). Feedback is delivered to both teachers and students; frequently, teachers receive feedback in the form of student-assessment results (Clark, 2012; Heritage, 2007; Heritage, 2010b), while students receive feedback from their teachers in the form of grades or other written or verbal evaluations. Both examples of feedback can be used to compare students' knowledge and/or task performance with class goals and standards (Butler & Winne, 1995; Heritage, 2007); in other words, teachers and students are collaborators in achieving classroom objectives, and feedback, regardless of the actor receiving it, is a valuable metric in appraising that collaboration. In addition to teachers (Black & Wiliam, 2009), peers (Sung et al., 2005), and computers (Clark, 2012), students themselves can generate feedback, in the form of internal feedback (Zimmerman, 2001). The defining criterion for feedback, regardless of its giver or recipient, is that it must be used for improvement; otherwise, it does not qualify as feedback (Black & Wiliam, 1998a).

Meta-analyses of previous research highlight the potential of feedback to enhance achievement, as long as the feedback meets certain criteria. A meta-analysis of 40 studies on the effect of feedback on achievement (Bangert-Drowns et al., 1991), mostly short-term studies at the college level, showed, on average, a positive effect of feedback on achievement. This positive effect was most manifest, however, when learners were provided guidance toward arriving at or were directly given the correct answers, as opposed to simply being told whether an answer is right or wrong. Moreover, feedback was most effective when it was delivered in a way that required learners to examine information and apply higher-order thinking skills, such as discriminating relevant information and using it to reflect, make new connections, and construct new learning. Another caveat about effective formative assessment was enunciated in Kluger and DeNisi's (1996) meta-analysis of 131 studies that examined the impact of feedback interventions (FI), "actions taken by (an) external agent to provide information regarding some aspect(s) of one's task performance" (p. 255) on achievement. While they found an overall positive impact of feedback on achievement, over 1/3 of the studies showed a negative impact, most commonly when the feedback, even when positive, targeted the student rather than the task.

A review of the FA literature reveals that, although researchers agree that FA is a tool that helps guide instruction and enhance student learning, there is variety in the exact way the concept is operationalized, especially in how it incorporates feedback. Some maintain, as Clark (2012) did in his framework, that feedback is a necessary part of FA, while others do not. For example, Harlen and James (1997) incorporated Ramaprasad's (1983) definition of feedback—"information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way" (p. 4)—into their own definition of FA—"essentially feed-back both to the teacher and to the pupil about present understanding and skill development in order to determine the way forward" (p. 369)—and thereby posited feedback to students as a requirement of FA. Black and Wiliam (1998b), however, asserted that FA occurs "when the evidence is actually used to adapt the teaching to meet student needs" (p. 140), which suggests that any assessment, whether feedback is provided directly to the students, is formative as long as its results are used to plan future instruction. Finally, Popham (2011) stated of FA, "This process revolves around the use of assessments to collect evidence, and then the employment of such evidence by teachers and/or students to decide whether they need to adjust what they are doing" (p. 36); this implies that FA may include feedback to students but is not the defining feature. In light of these discrepant formulations, for the purposes of this study, FA will represent the common thread in all of them: the broad concept of an assessment whose results are used to modify future learning experiences. While this study is grounded in Formative Assessment theory, FA, as demonstrated by the literature review, involves a vast number of components, and, in order to narrow the focus, will center around the concept of feedback and its effect on students' peer assessment self-assessment, and mathematics achievement.

Feedback. When providing feedback based on the results of collaborative discourse, teachers should ask themselves whether the feedback is crafted and delivered in a way that empowers students to take control over their own process of planning their subsequent goals and plans for meeting these goals (Davis & McGowen, 2007; Nicol & Macfarlane-Dick, 2006). Just as with metacognition, self-efficacy can be damaged by grades and other performance-oriented forms of judgment, such as comparison between students (Black and Wiliam, 1998a; Heritage, 2007; Nicol & Macfarlane-Dick, 2006), and learning can decrease (Black & Wiliam, 2009). The ideal is feedback based on effort and specific behaviors rather than achievement (Nicol & Macfarlane-Dick, 2006), but, if grades are adjudged absolutely necessary, they should be given only after students have received feedback in the form of specific comments (Gibbs & Simpson, 2004).

Even if the content and delivery of feedback are ideal, students' interpretations of feedback are shaped and can be distorted by factors such as their comfort level with a particular topic, previous educational experiences, motivation, or beliefs about themselves as learners (Clark, 2012; Nicol & Macfarlane-Dick, 2006). A learner's background can also affect whether feedback is accepted and applied (Butler & Winne, 1995). During classroom dialogue with students, teachers may uncover goals or beliefs that inhibit the positive student achievement that is a demonstrable result of well-applied feedback (Butler & Winne, 1995); an example of this is a student who believes it is more important to entertain her friends than it to learn. By extension, therefore, dialogue surrounding external feedback between students and external-feedback sources must be encouraged (Nicol & Macfarlane-Dick, 2006). To avoid miscommunication, it is recommended that, beyond providing opportunities to discuss feedback, teachers employ stimuli to engage students in dialogue about the feedback. Nicol and Macfarlane-Dick (2006) articulated specific dialogical feedback strategies:

> (a) providing feedback using one-minute papers in class; (b) reviewing feedback in tutorials wherein students are asked to read the comments they have been given on an assignment and to discuss these comments with peers (they might also be asked to suggest strategies to improve performance next time); (c) having students identify one or two examples of feedback comments that they found useful and explain why; (d) having students give each other descriptive feedback on their work in relation to criteria set before submission; and (e) group projects, especially those in which students discuss criteria and standards before the project begins. (p. 11)

Finally, engaging students in dialogue helps teachers detect when students misunderstand feedback and/or struggle to effectively absorb feedback into their plan for improved future performance (Nicol & Macfarlane-Dick, 2006). SRL is also enhanced when peers and students themselves are sources of feedback as they engage in peer and self-assessment (Perry et al., 2010).

As stated previously and reflected throughout, Clark (2012) provided a thorough theoretical background for the development of several components of the proposed intervention. Students' ability to accurately complete an algebra-based constructed response problem and their ability to accurately apply assessment criteria to the work of a peer and to their own work was assessed in a whole-group discussion format. Here, students' statements and answers to teachers' questions, which were crafted to elicit their thinking, generated data about their understandings and misconceptions that the teacher used to develop subsequent instructional sequences. The format of the lesson, an interactive whole-group discussion that valued the contributions and participation of all students, reflected the importance of practices that promote collaboration, discourse, and a respectful climate, all of which contribute to students' self-efficacy and subsequent achievement. Just as feedback is at the center of Clark's TFA, it was the crux of this intervention, because students used feedback provided by the teacher and their peers during the discussion as a tool to calibrate how accurately they applied assessment criteria to their own work and to that of a peer. The opportunity provided at the end of the sequence to make any adjustments they might deem necessary to their original assessments, based on feedback, promoted the mastery-goal focus that is recommended for formative experiences. And the relative ease of making adjustments to their assessment, through a short rubric, increased the likelihood that they will not conclude that the effort required to make adjustments outweighed the value of improving their previous work.

Intervention Components 2 and 3: Self- and Peer Assessment

Clark's (2012) TFA foregrounds the importance of student assessment in learning. The particular approach embraced by Clark makes students full partners in assessment and designs and encourages behaviors and activities toward that end.

Outlining some of these, Heritage (2007) and Phielix, Prins, Kirschner, Erkens, and Jaspers (2010) envisioned a classroom culture wherein self- and peer-assessment support were paramount.

Despite research that touts the benefits of self- and peer assessment in advancing achievement, in the minds of many teachers and students, these benefits are overshadowed by doubts about assessment accuracy. In order to promote the use of these strategies, this study sought to investigate practices that will close the gap between students' and teachers' assessments and thereby alleviate some of these doubts. The intervention is supported by Clark's (2012) TFA (see Figure 2.2 below); the components of the intervention will derive from Clark's foundational principles and the effective self- and peer-assessment strategies he proposed. The final section of this chapter will outline and recapitulate research that links the components of the intervention and the theorized effects on the outcome variables.

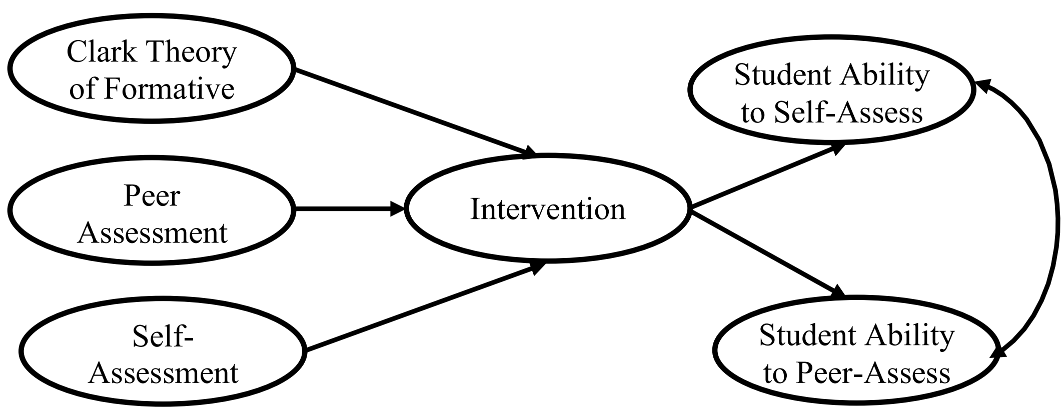


Figure 2.2. Clark's TFA and the proposed intervention. The application of the TFA, it is hypothesized, will yield reciprocal benefits for students' capacities to self- and peer-assess.

Explaining the discrepancy. While some teachers and even some students believe that students are dishonest—for example, teachers have raised concerns that "'mark sharks' will intentionally inflate their achievement, lying about their effort or misapplying the criteria," and students have speculated that "People could just take advantage of it and just mark all perfect when it's really not their best" (Ross, 2006, p. 8)—most students, even when their assessment is incorrect, are assessing honestly. Inaccurate assessments generally stem from student misunderstanding of criteria and how to apply it, not from intentional dishonesty (Black & Wiliam, 1998b). This notion is supported by the closer consistency between students' self-assessments and the assessments of their peers rather than between self- and teacher assessments (Ross, 2006), which, it can be argued, indicates that students share a likemindedness about what has been asked of them and how to execute it.

Teaching "evaluative skills." By sharpening their self-assessment skills, students are simultaneously growing their capacity for self-regulation. This stems from the coincidence between the actions involved, namely, comparing their own work with pre-established standards, predicting the degree to which they met predetermined criteria for success, and determining their level of satisfaction with how their work met the criteria (Ross, 2006). It has been noted in previous research that in order to accurately assess their performance against a predetermined set of standards, students "must already possess some of the evaluative skills as their teacher" (Nicol & Macfarlane-Dick, 2006, p. 3). Because not all students will arrive in class equipped with these skills, they must be explicitly taught how to carry out the process of effectively evaluating work (Andrade & Valtcheva, 2009).

Developing and communicating goals. As one student in a self-assessment reactions study was quoted, “You can self-assess all you want, but if you don’t know exactly what the teacher’s looking for, then how do you know?” (Andrade & Du, 2007, p. 171). The research is clear that students must begin the assessment process with an understanding of the learning goals and criteria; these must be clearly conveyed to them or developed with the students’ cooperation (Black & Wiliam, 2009; Ross, 2006; Thomas et al., 2011; Wiliam, 2007a). Once students have a clear understanding of the criteria, the teacher must model for them how to apply the criteria (Ross, 2006; Ross et al., 2002) and give them opportunities to apply it themselves (Ross, 2006). In order to examine the effectiveness of these strategies, I introduced them into the design of my study intervention. Students began each instructional session by completing a mathematics constructed response question, using as a guide rubric criteria that were provided in advance. Students then used the assessment criteria to assess their own work and a sample response to the same task, simulated to appear as that of a peer, Student X. Finally, I modeled for students how to apply the criteria to the same model student response and also requested student input during the process.

Peer and teacher feedback. The peer- and teacher-feedback component of the intervention emerged from Clark’s TFA and other previously cited literature attesting to the value of peer-assessment feedback (Andrade & Valtcheva, 2009; Clark, 2012; Davis & McGowen, 2007; Leahy et al., 2005; Nicol & Macfarlane-Dick, 2006; Ross et al., 2002; Sadler, 1989; Stiggins, 2009; Wiliam, 2007a). After I modeled assessment-criteria application and students were allowed to practice it themselves, they received ample feedback from peers and from me to help them gauge the quality of their assessments.

While implementing the intervention, I led 10 sessions of a self-assessment and peer-assessment program. At the beginning of each session, students independently completed a mathematical task. They then applied the assessment criteria to their own work and to that of an anonymous student, Student X. Upon completion, the class as a whole applied the criteria to Student X's work through discussion structured by peer-assessment feedback and teacher feedback. Students used this feedback to ascertain the relative accuracy of their own assessments.

Opportunities for revision. After the class performed this process, students reexamined their own individual self-assessments and assessments of Student X and made adjustments, if they believed any were necessary. The kind of feedback entailed for meaningful revision is evidence-based, specific, non-evaluative, directly related to the criteria, and developed over the course of class discussion, as led by the teacher (Butler & Winne, 1995; Clark, 2012; Nicol & Macfarlane-Dick, 2006). This opportunity for modifying work in order to improve it demonstrates to students the reason behind and the value of analyzing their work in the first place (Noonan & Duncan, 2005; Ross, 2006).

Growth over time and assessment reciprocity. As Figure 2.2 above illustrates, components based on self-assessment, peer assessment, and Clark's TFA theory form the proposed intervention. The intervention was premised on the conjecture that it will positively affect students' ability to self- and peer-assess; this conjecture derives from previous research that demonstrates that students' self-regulatory strategies are enhanced when students learn about and practice effective self-assessment skills (Boud & Falchikov, 1989). The goal was to assess if, as students became more accurate in their ability to self-assess (Black & Wiliam, 1998a), they would also become more self-directed and autonomous in their learning (Boud & Falchikov, 1989) and progress from assessment criteria that are highly structured for them ahead of time to less-defined criteria and even to develop their own criteria (Thomas et al., 2011). The process of peer assessing has been shown to, over time, cultivate students' ability to compare work against criteria and then transfer this skill into improved self-monitoring and revision of their own work (Nicol & Macfarlane-Dick, 2006); this study also sought evidence, found in the previous literature, that heightened self-assessment abilities allow students to better critique the work of others as well. This theorized reciprocal effect of self- and peer assessment on one another is displayed in the study framework.

Study Contributions to the Literature

The existing literature demonstrates a positive correlation between FA strategies, especially peer-assessment and self-assessment training, and student achievement. Unfortunately, the literature also exposes teachers' and students' doubts about the efficacy and integrity of student assessments. While these concerns are largely unfounded, they have been able to fester because most studies on FA are dated and used unclear constructs and questionable study design. In addition, most of the studies have focused on higher education students at the expense of their middle school counterparts. The next chapter of this study will describe an intervention that addresses these gaps.

CHAPTER III: METHODOLOGY

Restatement of Purpose and Research Questions

Recent data on mathematics achievement in the US (USED, 2013) indicate that many U.S. students are not performing at a level that will prepare them to successfully meet the demands of an increasingly global economy (NCTM, 2013a). Consequently, national leaders and organizations have advocated for changes in the U.S. education system; a well-publicized change was the development of the CCSS for K-12 English language arts and mathematics (NGACBP & CCSSO, 2010) that are designed to deliver curricula more commensurate with those of more-highly-achieving countries (Porter et al., 2011). Despite the prominence of these advances, the most significant opportunities for increased levels of learning occur in the day-to-day proceedings in mathematics classrooms (Black & Wiliam, 1998b; Noonan & Duncan, 2005). As established in the previous chapter, the potential benefits of formative assessment for all students, especially those who struggle (Black & Wiliam, 1998b), are well-documented (Andrade & Valtcheva, 2009; Brown & Hirschfeld, 2007; Davis & McGowen, 2007; Heritage & Heritage, 2013; Leahy et al., 2005; Nicol & Macfarlane-Dick, 2006; Ross et al., 2002; Suurtamm et al., 2010; Thomas et al., 2011; Wiliam, 2007b). Also prevalent in the research is the notion that embedding student self-assessment and peer assessment within a formative assessment plan yields positive outcomes for students (Heritage, 2007; Noonan & Duncan, 2005; Wiliam, 2007a).

Despite the manifold advantages of self-assessment and peer assessment, teachers sometimes hesitate to implement these strategies, fearing that, generally, students cannot accurately appraise their own work (Noonan & Duncan, 2005). However, strategies that allow students to more closely align their assessments with those of their teachers (Kulkarni et al., 2013; Lawson et al., 2012; Ross et al., 2002; Sadler & Good, 2006) do indeed exist. To this end, this study tested an intervention that combined certain of these strategies into an instructional sequence. The study sought to determine whether the intervention affected students' ability to accurately self-assess and peer-assess a mathematics task. In order to supplement the previous research that shows an effect of self- and peer assessment on academic achievement (Andrade & Valtcheva, 2009; Black & Wiliam, 1998a; Brown & Hirschfeld, 2007; Nicol & Macfarlane-Dick, 2006), this study also investigated whether the intervention produced a difference in middle school students' ability to successfully complete a mathematical task.

Population and Sample

Population

The population of this study consisted of 7th-grade students at an urban middle school in a metropolitan area in Kentucky. According to the 2013-2014 Kentucky School Report Card Data (2014a), more than 70% of the school's students received free or reduced-priced meals, and approximately 18% were classified as having special needs. The school has been designated as a Priority School, meaning that it has been "Persistently Low-Achieving" and "Needs Improvement." According to archived Kentucky School Report Card Data (2014b), from 2004 through 2011, the school did not make adequate yearly progress in mathematics; by 2011, it had fallen in overall performance to the 3rd percentile of all Kentucky middle schools. And while 2014 School Report Card data show that the school made adequate progress during the 2012 and 2013 school years, it failed to do so in 2014, when only 25.5% of students performed at proficient or distinguished levels on state mathematics tests, compared with 44.8% statewide. Given the aforementioned research findings that effective formative assessment helps close achievement gaps (Black & Wiliam, 1998a), the school was adjudged an appropriate location for the study.

Sampling and Assignment Procedures

Determining sample size. In conducting a power analysis to determine an adequate sample size, the researcher followed recommendations outlined by Field (2009). This entailed using the standard α level of .05, the recommended power of .8, and an estimate of the study's effect size. The researcher sought effect sizes in previous self- and peer-assessment research in order to anticipate the effect size appropriate for this study.

As previously noted, one of this study's objectives is to contribute to the relatively small amount of self- and peer-assessment research conducted in a middle school setting. Given this sparseness, in order to locate self- and peer-effect sizes to use as a basis for this research estimate, it was necessary to consider previous research that covered a variety of settings and populations. Falchikov and Boud's (1989) meta-analysis of studies examined self-assessment within a variety of contexts in higher education and provided 31 standardized mean difference statistics and 44 correlation coefficients from within the individual studies. Eva and Regehr (2011) reported the mean of six different correlations that compared psychology undergraduate students' predictions of how they would perform on a certain quiz with the students' actual performance. Ross (2006) reviewed previous self-assessment research across grade levels and content areas and reported effect sizes for self-assessment training and achievement. Comparisons between students' self-assessments and test scores focused on dental school exam results (Curtis, Lind, Dellinges, Setia, & Finzen, 2008) and graduate school applicants' prior SAT scores (Ross, 2006). Studies that compared teacher and student evaluations of student work and competency examined elementary school students' drawings (Butler, 1990), webpages developed by 14- and 15-year-old students (Sung et al., 2005), professional expertise of undergraduates in a department of social work (Baartman & Ruijs, 2011), essays written by undergraduates in a biology department (Kirby & Downs, 2007), and college students' course grades (Leach, 2012). Andrade et al. (2008) examined the effect of self-assessment training on elementary school students' proficiency in writing. Finally, Falchikov and Goldfinch's (2000) meta-analysis reviewed 48 studies that compared peer and teacher assessments in a variety of formats, including tests scores and behavioral ratings, across various higher education contexts. The study produced 24 standardized mean difference statistics and 55 correlation coefficients from within the individual studies. The list of correlation coefficients ranged from a minimum of $r = .02$ to a maximum of $r = .99$; due to the large range, the median $r = .5$ was calculated and applied. The standardized mean differences ranged from a minimum $d = 0$ to a maximum $d = 7.34$, so the median $d = .41$ was used. Because a d equal to .5 or an r equal to .3 indicates a medium effect size (Cohen, 1992), it was estimated that a medium effect size might be found in this study's data.

A number of other considerations entered into the process of determining an adequate sample. This study had an independent variable with two groups, treatment and no treatment, and three dependent variables, achievement, students' ability to accurately self-assess, and students' ability to accurately peer-assess. A covariate, previous achievement, which, research shows, influences a student's ability to assess accurately (Boud & Falchikov, 1989) was also considered. Because students' pre-assessment data for a particular study's measures may also be used as a covariate (Stevens, 2002), students were assessed before the intervention on their levels of achievement, ability to accurately self-assess, and ability to accurately peer assess. These measurements, which will be described later, were used as covariates as well. Given these factors, a multivariate analysis of covariance (MANCOVA) was deemed the most appropriate way to analyze the data (Mertler & Vannatta, 2002). To determine an adequate sample, the researcher took the standard α level of .05, the recommended power of .8, the medium estimated effect size, the fact that a MANCOVA would be used to analyze the data, and a power analysis, and arrived at a desired sample size of 48 (Statistics Solutions, 2014).

Use of stratification. Previous research has found a relationship between students' achievement and their ability to accurately assess (Boud and Falchikov, 1989; Falchikov & Boud, 1989). As indicated earlier, only 25% of students school-wide at the study site had demonstrated proficiency in mathematics during the previous school year. Therefore, a random sample could have included very few or no high-achieving students. In order to test the intervention across achievement levels, a stratified sample was deemed appropriate (Teddlie & Yu, 2007) to ensure that the sample would cover the complete range of previous student achievement (Creswell, 2009). Academic achievement was measured using the most recent Measures of Academic Progress (MAP) assessment (Northwest Evaluation Association [NEA], 2012) that is administered to every student in the school three times a year. For the purposes of this study, student achievement was divided into three categories: low, medium, and high. A “high” categorization was determined by the likelihood, based on research provided by the developers of the MAP assessment (NEA, 2013), that a student would perform at a proficient or distinguished level on the Kentucky state assessment; a “medium” categorization was determined by the likelihood that a student would perform at an apprentice level on the state assessment; and a “low” categorization was determined by the likelihood that a student would perform at a novice level on the state assessment. According to the aforementioned report card data, fewer than 4% of students at the study site displayed distinguished levels of achievement in the previous year, and school-wide MAP assessment data indicated that less than 2% of students showed such levels. School-wide state report card data revealed that about 21% of students performed at a proficient level in mathematics, while MAP data reflected a total of about 15% of students. In order to best represent the range of ability levels typical of the study site, students who had displayed both distinguished and proficient achievement levels on the MAP assessment were assigned to the “high” category, because, relative to most students, they were high achieving.

Sampling scheme. This study’s sampling scheme contained elements of convenience sampling and purposive sampling, as the process moved from the school to an individual student level. The study site was chosen for convenience (Teddlie & Yu, 2007): The researcher’s teaching position there provided ready access, and the school’s administration was willing to participate in the study. Because individual and classroom sources of variance and correlation impact outcome variables in educational research

(Koth, Bradshaw, & Leaf, 2008; Muthén, 1990; Ryan & Patrick, 2001), this study sample was drawn from an intentionally narrow portion of the school's population, reflecting a purposive selection method (Teddlie & Tashakkori, 2009). In order to minimize the number of classroom factors that could influence the intervention results, students from only 2 seventh-grade classroom teachers—henceforth, Teacher A and Teacher B—were chosen for the study's treatment and control groups, with the goal of obtaining an equal number of treatment and control group students from each teacher's classes.

Individual student selection and assignment to treatment and control groups. The MAP measure of previous mathematics achievement was collected for all students of both teachers and was used, in part, to determine students' assignment to treatment and control groups, in a purposive manner (Teddlie & Tashakkori, 2009). The goal was to create treatment and control groups drawn equally from each teacher and having comparable levels of previous achievement. Of Teacher A's students, 6 displayed high levels of achievement; therefore, in order to form control and treatment groups of comparable size, 3 students were to be assigned to the treatment group and 3 were to be assigned to the control group. Twenty-two students displayed a medium level of achievement; therefore, 11 each were assigned to the treatment and control groups. Finally, 11 students who displayed a low level of achievement were assigned to fill the treatment group, and 11 were selected to fill the control group. Of Teacher B's students, 8 displayed high levels of achievement; therefore, in order to form control and treatment groups of comparable size, 4 were assigned to the treatment group and 4 were assigned to the control group. Twenty-six students displayed a medium level of achievement; therefore, and 11 each were assigned to the treatment and control groups. Finally, 10 students who displayed a low level of achievement were selected to fill the treatment group, and 10 were selected to fill the control group.

Once levels of previous mathematics achievement within the seventh-grade population's students were determined, the risk of bias during their assignment to treatment or control groups could have been lessened by random selection and assignment to treatment or control groups (Shadish, Cook, & Campbell, 2002). However, random assignment was not feasible for this study because certain constraints of the real-life education setting compelled the researcher to make some of the assignments strictly out of convenience. Assessment instruments and the intervention were administered during an intervention and enrichment class that all students attended daily. Students lost no core content class instruction from participating in the study intervention. Moving students into and out of classes required the school's counseling office staff members to make relatively cumbersome schedule changes, and their other responsibilities prohibited them from accommodating a large number. Therefore, treatment and control assignments were made to minimize the number of necessary class changes. At the study site, seventh-grade students are routinely assigned to an intervention or enrichment class led by one of the four teachers on their core content team. In choosing a treatment group, prior to the beginning of the study, the researcher examined the rosters of the four intervention and enrichment classes already in place on each of the two core-content teams. The class on each of the two teams whose student makeup most closely resembled the needed treatment group, as developed from previous mathematics MAP achievement data, was selected to receive the treatment. A small number of class changes were then made to create a class with the desired number of students at each achievement level. These schedule changes were also dictated by factors related to students' needs as judged by the team teachers. For example, 40 students with novice levels of previous mathematics achievement, who therefore could have been assigned to the treatment groups, were required by their language arts teachers to remain in their language arts intervention class, and so could not participate in the study's mathematics-oriented intervention. Most of these 40 students were permitted to be in the control group, which required students to undergo just two intervention sessions, one before the treatment and one after participating in the assessments; however, nine of those students were held back from participating in any way. Significantly, the approach by which the school organizes classes rendered the desired sample of 48 students impossible. The school's need to balance enrollment in intervention and enrichment classes meant that each intervention and enrichment class must consist of approximately 25 students. Therefore, the two classes that received the study intervention, one class from each teacher, were planned to consist of 25 students each. To have equal distribution of students in treatment and control groups, the researcher sought to include 25 students from each seventh-grade mathematics teacher in the control group. Given the concerns about the potential low power of the study, the higher sample size entailed by the study-site factors could enhance the study's power. It was hoped that each of the two teachers could contribute 50 students, 25 for the treatment group and 25 for the control group, for a total sample of 100 students.

The target sample of 100 students was not met, due to some schedule changes dictated by students' needs, student absenteeism throughout the intervention period, and absenteeism on control group pre- and posttest administration days. These factors yielded

39 students in the treatment group, 18 in Teacher A's class, and 21 in Teacher B's class. The control group consisted of 41 students, 20 from Teacher A's mathematics class and 21 from Teacher B's class. Despite a smaller-than-planned sample, the total of 80 still significantly exceeded the minimum number of 48 participants required by the power analysis (Statistics Solutions, 2014). Table 3.1 below indicates the number of subjects in each category.

Table 3.1

Number of Students in Treatment and Control Sections for Teachers A and B

Teacher	**Treatment**				**Control**				
	High	Med	Low	Tot	High	Med	Low	Tot	Total
A	3	8	7	18	3	7	10	20	38
B	4	7	10	21	3	8	10	21	42
Total	7	15	17	39	6	15	20	41	80

Research Design

Type of Research

The type of research employed in this study may be classified as quantitative, which Creswell (2009) defined as "means for testing objective theories by examining the relationship among variables. These variables can, in turn, be measured, typically on instruments, so that numbered data can be analyzed using statistical procedures" (p. 4). The independent variable of this study was involvement in the treatment intervention versus non-involvement, and the dependent variables were students' ability to accurately assess their own work, their ability to accurately peer-assess, and their mathematics achievement. Because all of the data collected in this study were numeric and analyzed using a statistical procedure, quantitative research, based on the previous definition, was adjudged the appropriate methodology (Teddlie & Tashakkori, 2009). The study design, including variables, measures, and instruments, is elaborated in Table 3.2 and Table 3.3.

Table 3.2

Study Design

Group	**Pretest Variables**	**Intervention**	**Posttest Variables**
NR Treatment N = 39	O1 O2 O3 O4	X	O5 O6 O7
NR Control N = 41	O1 O2 O3 O4		O5 O6 O7

Table 3.3

Variables, Measures, and Instruments

Variable	**Variable Measure**	**Instrument**
O1 Covariate 1 Previous Academic Achievement (NAPD)	Determined by Spring Mathematics MAP scores from 2015	MAP test (NEA, 2013)

O2 Covariate 2 Peer-Assessment-Accuracy Pretest (PREPEER)	Determined by absolute value of discrepancy between peer and teacher assessment, based on assessment item rubric	Assessment Item and Teacher/Peer Rubric
O3 Covariate 3 Self-Assessment-Accuracy Pretest (PRESELF)	Determined by absolute value of discrepancy between student and teacher assessment, based on assessment item rubric	Assessment Item and Self-Assessment Rubric
O4 Covariate 4 Achievement Pretest (PREACH)	Determined by score obtained, based on assessment item rubric	Assessment Item and Teacher/Peer Rubric
O5 Peer-Assessment-Accuracy Posttest (POSTPEER)	Determined by absolute value of discrepancy between peer and teacher assessment, based on assessment item rubric	Assessment Item and Rubric
O6 Self-Assessment-Accuracy Posttest (POSTSELF)	Determined by absolute value of discrepancy between student and teacher assessment, based on assessment item rubric	Assessment Item and Self-Assessment Rubric
O7 Achievement Posttest (POSTACH)	Determined by score obtained, based on assessment-item rubric	Assessment Item and Teacher/Peer Rubric

Variable Measurement

All mathematical tasks and the rubrics used to assess them were designed by the researcher; the design principles and process will be detailed in the instrumentation section.

At the beginning of the study, students in the treatment and control groups completed the first mathematical task, displayed below in Figure 3.3. They subsequently assigned their work a score within a 0-12 range by applying provided rubric criteria, seen here in Figure 3.1.

Self-Assessment Rubric - Multiple Choice Question *Directions: Place a checkmark in the blank next to each indicator that your answer has fulfilled* Correct Answer Selection/Explanation	Incorrect Answer Selection/Explanation
____ 1 Point: I am certain I identified the correct multiple choice answer OR ____ 1 Point: I think I probably identified the correct answer, but I'm not certain	____ 1 Point: I am certain I identified an incorrect multiple choice answer OR ____ 1 Point: I think I probably identified an incorrect answer, but I'm not certain.

<table>
<tr><td>____ 1 Point: I explained my answer with some correct words, mathematical symbols, pictures, or other evidence.

OR

____ 2 Points: I completely and clearly explained my answer with correct words, mathematical symbols, pictures or other evidence.</td><td>____ 1 Point: I explained why the answer was incorrect with some correct words, mathematical symbols, pictures, or other evidence.

OR

____ 2 Points: I completely and clearly explained why the answer was incorrect with correct words, mathematical symbols, pictures or other evidence</td></tr>
<tr><td colspan="2">Self- Assessment Rubric - Constructed Response Question
Directions: Place a checkmark in the blank next to each indicator that your answer has fulfilled

Part A Part B</td></tr>
<tr><td>____ 1 Point: I completed every part of the question in Section A</td><td>____ 1 Point: I completed every part of the question in Section B</td></tr>
</table>

____1 Point: I correctly answered every part of the question in Section A	____1 Point: I correctly answered every part of the question in Section B
____ 1 Point: I clearly explained every part of my answer in Section A with correct words, mathematical symbols, pictures or other evidence.	____ 1 Point: I clearly explained every part of my answer in Section B with correct words, mathematical symbols, pictures or other evidence.

Figure 3.1. Student Self-Assessment Rubric sample. Students were charged with assigning a score that expressed their perceived success or failure in satisfying the stated conditions.

The students subsequently applied the criteria, offered here in Figure 3.2, to score another answer to the same task, one that the researcher had simulated to resemble that of a student peer.

Teacher/Peer Assessment Rubric - Multiple Choice Question	
Directions: Place a checkmark in the blank next to each indicator that the student's answer has fulfilled	
Correct Answer Selection/Explanation	Incorrect Answer Selection/Explanation

____ 1 Point: THE STUDENT identified the correct multiple choice answer	____ 1 Point: THE STUDENT identified an incorrect multiple choice answer
____ 1 Point: THE STUDENT explained his/her answer with some correct words, mathematical symbols, pictures, or other evidence. OR ____ 2 Points: THE STUDENT completely and clearly explained his/her answer with correct words, mathematical symbols, pictures or other evidence.	____ 1 Point: THE STUDENT explained why the answer was incorrect with some correct words, mathematical symbols, pictures, or other evidence. OR ____ 2 Points: THE STUDENT completely and clearly explained why the answer was incorrect with correct words, mathematical symbols, pictures or other evidence

Teacher/Peer Assessment Rubric - Constructed Response Question *Directions: Place a checkmark in the blank next to each indicator that the student's answer has fulfilled* Part A	Part B
____ 1 Point: THE STUDENT completed every part of the question in Section A	____ 1 Point: THE STUDENT completed every part of the question in Section B
____1 Point: THE STUDENT correctly answered every part of the question in Section A	____1 Point: THE STUDENT correctly answered every part of the question in Section B
____ 1 Point: THE STUDENT clearly explained every part of his/her answer in Section A with correct words, mathematical symbols, pictures or other evidence.	____ 1 Point: THE STUDENT clearly explained every part of his/her answer in Section B with correct words, mathematical symbols, pictures or other evidence.

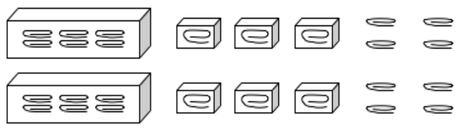
Figure 3.2. Teacher-/Peer-Assessment Rubric sample. This rubric was designed to generate comparisons in assessment capabilities.

The researcher then applied the teacher/peer rubric criteria to each student's completed task; the researcher's score was used to quantify students' pretest mathematical achievement score. The absolute value of the discrepancy between the students' self-assessment scores and the researcher's scores was calculated and used as the study's pretest scores to measure students' ability to accurately self-assess. For example, if Student A scored her task a 3 and the researcher scored it a 4, the self-assessment discrepancy score was a 1. Before the study began, the researcher applied the aforementioned teacher/peer rubric to the simulated peer response and obtained a score of 8. The absolute value of the discrepancy between the students' peer-assessment scores and the researcher's scores was calculated and used as the study's pretest scores to measure their ability to accurately peer-assess. For example, if Student B scored the simulated peer response a 10, her peer-assessment discrepancy score was a 2. Two pretest scores were collected using this approach for all 80 treatment and control students in the study. The same method was used to collect posttest data for the treatment and control groups, as the students completed, self-assessed, and peer-assessed the same task.

Instrumentation

Description of instruments. This study's instrumentation comprised the aforementioned assessment, the self-assessment rubric, and the peer-/teacher-assessment rubric. A number of factors helped enable the researcher to minimize the potential for measurement error and to thereby increase the likelihood that these tools indeed measure students' ability to assess and their mathematics achievement—in other words, to maximize their reliability and validity (Kimberlin & Winterstein, 2008).

The assessment used for the pre- and post-assessment consisted of one multiple choice question and one constructed response question from the Diagnostic Teacher Assessment of Mathematics and Science (DTAMS) Instrument (Saderholm, Ronau, Brown, & Collins, 2010). An example of these questions is displayed in Figure 3.3 below.



A boy has 2 large boxes of paper clips, 6 small boxes of paper clips, and 8 individual paper clips. Using L to represent the number of paper clips in each large box and S to represent the number of paper clips in each small box, write an expression to represent the total number T of paper clips that he has.

a. $2L + 6S + 8$
b. $L + S + 6$
c. $2S + 6L + 8$
d. $2L \times 6S \times 8$

*** Write the correct answer choice here _________**
***Explain, using any words, diagrams, math symbols, or other evidence you would like, why you think the answer is correct.**

***Choose one of the incorrect answer choices _____, and explain, using any words, diagrams, math symbols, or other evidence you would like, why you think the answer is incorrect.**

You received a DVD player for your birthday and have begun to compare the cost of renting DVDs at nearby video stores. Below are the prices charged at two video stores.

- Movies To Go: $5.00 per DVD
- Movies Are Us: A one time charge of $10 plus $4.00 per DVD

a. Write two equations that represent this situation.

b. Solve these equations and explain which store you might choose, and what circumstances would influence your decisions. Justify your selection.

Figure 3.3. Pre- and post-assessment items. The mathematics problems were drawn from the DTAMS instrument.

The content of the assessment was developed and evaluated by a team of mathematicians, mathematics educators, and middle school teachers, based on a literature review of middle school teacher and student misconceptions; analysis found the assessment items to be valid and reliable measures of teachers' content knowledge for middle school mathematics. Although the assessment was developed with teachers in mind, the researcher chose the two assessment items used in this study after a determination that they also aligned with content knowledge relevant to seventh-grade mathematics students.

Assessment-instrument validity. Several additional steps were taken to ensure the assessment item's appropriateness. Following the recommendations of McMillan (2001a), three major categories of validity evidence were sought: content-related evidence, criterion-related evidence, and construct-related evidence.

Evidence categories.

Content-related evidence. The researcher heeded McMillan's (2001a) recommendations for pursuing content validity, that is, evidence that the assessment measures students' mastery of an aligned standard and is consistent with the objectives of a seventh-grade mathematics course (Teddlie & Tashakkori, 2009). The researcher sought evidence of face validity and instructional validity with the assistance of experts familiar with the mathematics curriculum and sample students' instruction. The sample students' teachers reviewed the assessment instrument and found it to be consistent with their class content and instruction. The assessment was also reviewed by three other

National Board Certified mathematics teachers. They examined the items and the seventh-grade standards to determine whether there was evidence of face validity, and all agreed that the items aligned with Common Core standard 7.EE.B.4 (NGACBP & CCSSO, 2010).

Construct-related evidence. Once content validity was established, the items were examined for construct-related evidence of validity (McMillan, 2001a). In other words, an assessment intended to evaluate students' mastery of a certain algebraic-thinking standard must not test other constructs, such as students' reading ability or background knowledge of a topic, that can impede students' ability to demonstrate mastery of the standard. Construct-related evidence of validity is consistent with Moskal (2003), who suggested a number of criteria to consider when developing an assessment task, including the exclusion of any expectation that students demonstrate knowledge extraneous to the intended focus area of the assessment. The task in this intervention involved paper clips and DVDs and demanded no background knowledge outside the assumed typical life experience of a middle-school-aged student.

Criterion-related evidence. In order to determine whether an assessment actually measures a construct the researcher intends it to, one might investigate the correlation between the assessment measurement and another measurement that is likely a gauge for the same construct (Kimberlin & Winterstein, 2008). In this case, before the assessment score itself could be compared with another potential measure, the rubric that was used to calculate the assessment score needed to be analyzed.

Assessment rubrics. The Teacher-/Peer-Assessment Rubric and Self-Assessment Rubric are two rubrics with the same indicators of successful task-completion, each worded to suit their different audiences. The researcher created the rubrics with the input from three other experienced middle school mathematics teachers and one mathematics teacher educator. The rubric features a combination of analytic components, based on indicators of correct and complete answers to parts of the question, and a holistic component, based on students' overall demonstration of conceptual understanding, application, and communication (Moskal, 2000).

Instrument reliability. For the assessment rubrics to best evidence validity, they must be reliable, in that the scores they yield must be consistent with one another (McMillan, 2001b). The reliability of the rubric was examined during a field test, before it was used to investigate the validity of the assessment. The field test involved 48 students who were seventh graders at the same study site but not participating in the study. Once the field-test students completed the assessment task, a group of six middle school mathematics teachers from the study site gathered to score them according to rubric criteria. Each student's completed task was scored by two teachers using the Teacher/Peer Rubric. The degree of reliability was determined by calculating Cronbach's alpha, and a sufficient benchmark of reliability was found, $\alpha = .87$ (Jonsson & Svingby, 2007). Once the scoring was completed, the six teachers discussed the nine pairs of scores that differed by more than 1 point to determine the source of the discrepancy. It was found that in all but two cases, the discrepancies were the result of a teacher's misunderstanding of the rubric criteria, resulting in higher scores than would otherwise have been given. Once the criteria were clarified for this teacher, some of the scores were revised; the new, higher, Cronbach's alpha was calculated as $\alpha = .94$.

Criterion-related validity. Once the field-test teacher rubric scores were found to be reliable, they were used to obtain evidence of the assessment's criterion-related validity, by investigating the correlation between the study assessment measurement and another measurement that is likely a gauge for the same construct, mathematics achievement (Kimberlin & Winterstein, 2008). For the purposes of this study, if there was a discrepancy in the two field test teacher-calculated scores assigned to a single student, that student's achievement was calculated by taking the average of the two teacher scores. The researcher then compared the field test students' performance on the assessment task with their performance on a previous assessment that measured mathematics achievement (Chatterji, 2003), the mathematics scale score from a recent administration of the previously described MAP test (NEA, 2012). Because each of the two measures used a different scale, the researcher applied SPSS to obtain a Pearson correlation coefficient (Field, 2009). A correlation was found between the two variables [$r = .51$, $n = 80$, $p < .001$]. Because this correlation was significant and met the benchmark for a large effect size, .5 (Cohen, 1992), evidence of criterion-related validity was established for the assessment item.

Administering instruments. Treatment- and control-group students were given the pre- and post-assessments during academic intervention and enrichment time, granted daily to all students during school hours. Pre- and post-assessments had the same structure. Students completed the mathematical task using rubric criteria that had been provided ahead of time as a guide. They then used the rubric to assess their own task and the sample work of a simulated peer response to the same constructed response task. In order to ensure that the nature and purpose of the tasks were presented consistently throughout, the researcher was the only individual to administer the pre- and posttests.

Scoring instruments. After students completed the task, they applied the assessment-rubric criteria and assigned their work a score based on the rubric range. They were then presented with a simulated peer response to the same task. The researcher subsequently applied the same rubric criteria to each student's work and assigned a score both to each student's task response and to the simulated student work. The discrepancy scores were calculated from each student's self- and peer assessment of each task.

Intervention

As discussed in the previous chapter of this study, the intervention was influenced by strategies outlined in the FA literature, particularly that which centered on self-assessment and peer assessment. This section of the current chapter highlights those influences. A more in-depth description of the intervention plan, including specific learning goals, questions asked, and mathematical tasks used throughout, may be found in Appendix A of this study.

Intervention Structure

The instructional sequence occurred over 10 hour-long sessions across 3 weeks. In order to minimize potential unreliability of treatment implementation (Shadish et al., 2002), the researcher conducted all intervention sessions. The intervention was delivered on Mondays, Wednesdays, and Fridays during the first 2 weeks, and then Monday through Thursday during the 3rd week. Following recommendations from previous studies, the researcher engaged students in a variety of activities (Ames, 1992) that featured individual reflection, pair and whole-group discussion, and technology.

Setting a purpose and communicating learning goals. As noted in the previous chapter, Clark (2012) asserted the importance of considering students' personal factors and background knowledge and experiences when designing learning and instructional sequences. As the mathematics department chair at the study site, the researcher communicated regularly with math teachers about their classroom practices; these conversations revealed that the students involved in this study entered with little or no exposure to peer and self-assessment, at least for the two years they had spent at the study site. Although students' involvement in the assessment process can promote engagement, buy-in (Ross, 2006), and positive feelings toward assessment (Black & Wiliam, 1998a), students' inexperience with the assessment process could affect their willingness to engage throughout the intervention. Students more actively participate in tasks when they believe the tasks are relevant to them (Brophy, Rohrkemper, Rashid, & Goldberger, 1983); therefore, Session 1 of the intervention sequence was designed to give students the opportunity to find and develop a purpose in the assessment-training program as they reflected on and discussed their own previous experiences with assessment and how honing their self- and peer-assessment skills could improve their future assessment outcomes.

Communication of goals and criteria for success. Established in the previous literature review is the claim that, in order to effectively engage in the assessment process, students must enter it with an understanding of the learning goals and criteria; this understanding can be developed by simply conveying goals and criteria to students or by developing goals with the students' cooperation (Black & Wiliam, 2009; Ross, 2006; Thomas et al., 2011; Wiliam, 2007a). The case has been made, however, that the latter technique, including students in criteria development, best helps them develop a more thorough understanding of goals and criteria (Andrade & Du, 2007; Heritage, 2007; Nicol & Macfarlane-Dick, 2006; Ross et al., 2002). Both strategies were employed in Sessions 2 and 3 of the intervention.

At the beginning of Session 2, students were given the constructed response portion of the Teacher-/Peer-Assessment Rubric and a blank constructed response task. Student pairs created lists of possible characteristics for an assessment answer to satisfy each rubric indicator, then presented their ideas in a whole-group setting; the researcher recorded the ideas and assembled them into a master list. Through this process, two goals of the intervention were met. First, developed by mathematics educators, the rubrics shown above equipped students with indicators of successful task completion, which, due to common student misunderstandings of what good work looks like (Ross, 2006), they might not have developed on their own. Second, student input into the development of successful-task-completion characteristics accorded students the aforementioned benefits, including heightened enthusiasm confidence, and ownership, that accompany their participating in criteria development. Moreover, because the master list was developed in a whole-group setting, the researcher was able to guide students toward identifying and adding to their list any exemplary task characteristics they might have overlooked due to their lack of expertise.

In Sessions 5 and 6, students also developed a master list of successful-task-completion characteristics, but did so differently from before. Student pairs were first presented with the Self-Assessment Rubric and blank constructed response tasks, which they completed using the rubric as a guide. The whole group then reviewed each pair's response and, from these specific work samples, compiled evidence of successful task completion to create a master list of characteristics found in an exemplary response.

Opportunities to practice applying criteria. Previous research has uncovered the value of modeling for students how to apply assessment criteria (Ross, 2006; Ross et al., 2002) and of giving them opportunities to apply criteria themselves (Ross, 2006). These strategies were included throughout every session, after session one. Some sessions, they applied criteria to peer work, and, at other times, they applied criteria to their own response.

In Sessions 2 and 3, once students developed a whole-group master list of successful task characteristics for a blank task provided, they applied the master list, developed with the provided-teacher peer rubric in mind, to simulated peer responses and constructed a whole-class evaluation of the responses. As they engaged in this process, the researcher facilitated and modeled, when needed, how to apply criteria. In Session 4, students applied the Teacher/Peer Rubric to two more simulated student tasks, again with teacher guidance, to generate a group evaluation of successful task completion. The practice of applying criteria to peer work has been shown to be particularly valuable, because it allows students to practice and therefore sharpen their evaluation skills without the negative emotions that can sometimes come while analyzing one's own product (Wiliam, 2007a).

After three sessions of applying criteria to the work of others, students began generating and assessing their own work. In Sessions 5 through 10, students generated their own responses to mathematical tasks and applied the Self-Assessment Rubric as a guide. In Sessions 5 and 6, student-pair-generated assessment responses were subsequently reviewed as a whole group, as the class engaged, again, in the peer-assessment process and used rubric criteria to identify strengths and opportunities for improvement in each pair's responses. In Sessions 7 and 8, student pairs completed a task while applying self-assessment criteria; they then evaluated their own work against exemplar work samples (Orsmond, Merry, & Reiling, 2002) by citing specific evidence of similarities and differences between their own work and the provided exemplar samples. In Sessions 9 and 10, students completed their assigned tasks individually while applying self-assessment criteria, then were given simulated peer, non-exemplar, work samples for the same task; these were evaluated in a whole-group setting.

Peer and teacher feedback. Previous research attests that as students learn to participate in the assessment process, they must receive ample feedback from peers and their teacher to help them gauge the quality of their assessments (Andrade & Valtcheva, 2009; Black & Wiliam, 1998b; Clark, 2012; Davis & McGowen, 2007; Leahy et al., 2005; Nicol & Macfarlane-Dick, 2006; Ross et al., 2002; Sadler, 1989; Stiggins, 2009; Wiliam, 2007a). Various sources of feedback on students' assessment were integrated into Sessions 2 through 10.

Sessions 2 through 6, 9, and 10 provided opportunities for peer and teacher feedback through the earlier whole-group evaluation sessions. In line with recommendations from previous research, the researcher was careful to guide the discussions in a way that promoted feedback, with a task- and evidence-based focus, and the researcher contributed to the discussion only when necessary, ensuring that it would have a student-centered, peer-assessment focus. This emphasis on student voice, as opposed to that of a teacher "expert" alone, frequently leads to elevated discourse that would not occur solely with teacher feedback (Gielen et al., 2010).

Students were the recipients of written teacher feedback throughout the intervention, as well. Before engaging in Tasks 5, 8, and 10, students received specific, work-product-centered, written teacher feedback on their comparisons of their own work with peer work samples and on possible revision to their work based on these comparisons and the aforementioned sources of feedback. In Sessions 7 and 8, students received another kind of written feedback in the form of exemplary simulated peer responses (Orsmond et al., 2002).

Opportunities for revision. Previous research arms educators with many reasons to allow students opportunities to revise and resubmit work (Boud, 2000; Clark, 2012; Guskey, 2003; Nicol & Macfarlane-Dick 2006; Noonan & Duncan, 2005; Ross, 2006; Sadler, 1989; Stiggins, 2002; Zimmerman, 2001), including heightened student buy-in to a mastery focus on learning and motivation to improve work. Therefore, the researcher underscored the value of revision and improvement throughout the intervention. In Sessions 3 through 8, student pairs revisited their original assessments of simulated peer tasks, their own task responses, and self-assessments, as they compared and contrasted their own products with those generated by the group. In Sessions 9 and 10, they engaged in the same kind of analysis, but focused on work products that they had originally developed independently. These opportunities were always presented in the context of an end-goal of producing future work of improved quality.

Data

Collection Procedures

Rubric scores were recorded by the students and by the researcher on provided rubrics; the achievement and discrepancy scores were then calculated and entered into Microsoft Excel. Subsequently, all data were entered into statistical analysis software.

Statistical Analysis

Rationale for statistical procedure. A number of factors influenced the statistical procedure used in this study, including the number of variables and the presence of potential covariates. This study had an independent variable with two groups, treatment and no treatment, and three dependent variables, achievement, students' ability to accurately self-assess, and students' ability to accurately peer-assess. In addition, as mentioned repeatedly, previous achievement is potentially linked to students' ability to assess accurately (Boud & Falchikov, 1989; Falchikov & Boud, 1989), as is pre-assessment data for a study's particular measures (Stevens, 2002); therefore, students' previous performance on the MAP test and pretest scores on the study's measures of achievement, their ability to pre-asses, and their ability to self-assess were collected. In order to account for potential variance within the dependent variables' data that might be explained by these covariates, a three-level multivariate analysis of covariance (MANCOVA) was deemed appropriate (Field, 2009; Mertler & Vannatta, 2002; Stevens, 2002) and conducted in SPSS (Field, 2009).

Descriptive and inferential tests reported. To confirm that an analysis of covariance was indeed appropriate, correlation coefficients were calculated and reported. Dependent variable posttest and all covariate means and standard deviations were calculated for treatment and control groups and displayed in a table. Post-achievement means and standard deviations were also calculated for students based on their levels of previous achievement and displayed in a table. Wilk's Lambda *F* ratios, *p* values, and effect sizes *r* were calculated for the multivariate tests.

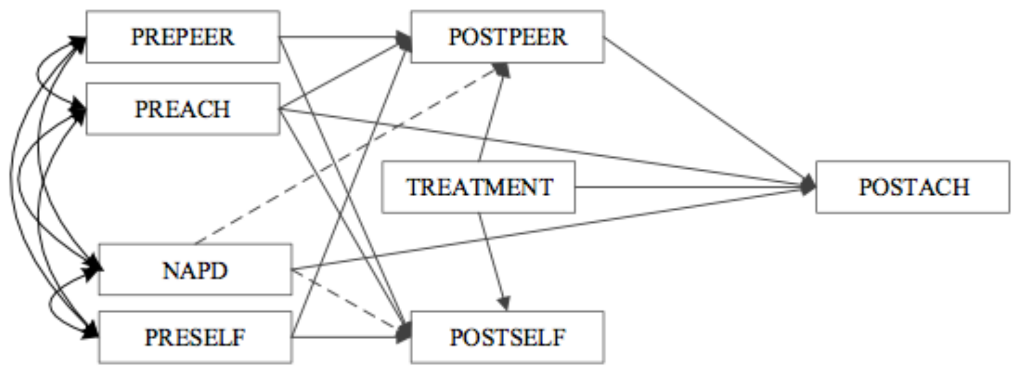


Figure 3.4. Study framework. The reciprocal interactions between the variables are depicted.

Theoretical Premise

The intervention was premised on the conjecture that it would positively affect students' ability to self- and peer-assess (POSTSELF) and (POSTPEER) and would, consequently, positively impact students' mathematics achievement (POSTACH). The researcher also posited a reciprocal effect of self- and peer-assessment ability. Due to the aforementioned links among students' previous achievement, current achievement, and ability to accurately assess, the covariate variables, PREACH, PRESELF, PREPEER, and NAPD, served as controls for previous student-ability levels.

Study Limitations

The sampling scheme outlined earlier highlights the most obvious limitation of this study. The sample consisted only of 7th-grade students from the mathematics classes of two participating teachers in an urban Kentucky middle school. Of all the students in those teachers' classes, study participants were selected and assigned to groups according to previous mathematics achievement and to school-related scheduling constraints on placing them in a mathematics intervention and enrichment class. Another possible limitation is the researcher's relative inexperience as a teacher with implementing a systematic series of self- and peer-assessment strategies within the course of one intervention. Moreover, self- and peer assessment were novel concepts for the students involved as well. Other limitations could include the study's short duration given the complexity of the intervention, the relatively small range of the instrument scores, and the reliability and validity of the assessment scores.

Validity Threats

Certain aspects of this study could indicate potential internal validity threats. The most apparent of these is the selection bias that could have occurred from the sampling scheme. It has also been noted that students' previous interaction with an assessment could have yielded in increased levels of achievement from the previous exposure alone (Shadish et al., 2002). The use of the same pre- and post-assessment instrument leaves this study open to this critique.

Subject selection and group assignment. This study's use of matching and purposive assignment for control and treatment groups could have introduced potential sources of bias (DiPrete & Gangl, 2004). Because the researcher had familiarity with this group of students, the researcher could have intentionally selected students for each group as to promote the likelihood the study hypothesis was found to be true. The researcher attempted to minimize other forms of bias, however, by having comparable numbers of students in the treatment and control group from each teacher and comparable numbers in each group of students drawn from accelerated classes, a population linked to superior assessment skills (Boud & Falchikov, 1989; Falchikov & Boud, 1989).

Construct validity. To fully explicate the constructs *mathematics achievement*, *peer-assessment accuracy*, and *self-assessment accuracy*, detailed descriptions of the tools used to measure these constructs have been provided, as have reports of steps taken to ascertain the reliability and validity of the assessment items and rubrics. The accuracy of students' assessments has been measured by the discrepancy between a teacher's score and a student's score; variance among teachers' scores when assessing the same pieces of student work could contribute to additional error variance in students' scores. According to previous research (Falchikov & Boud, 1989; Kulkarni et al., 2013; Sadler & Good, 2006), however, student-teacher discrepancy is, in the field, a commonly accepted measure of students' accuracy.

External validity. To minimize factors outside the intervention that could produce statistical variance, the setting of the study was very limited. Therefore, the results that follow are not generalizable to settings outside of this particular set of urban-middle school 7th-grade mathematics students.

CHAPTER IV: RESULTS

Introduction

This study sought to determine the effect of an intervention of self- and peer-assessment training for middle school students on their ability to accurately self-assess and peer-assess and on their mathematics achievement. Previous research has shown that self- and peer-assessment training can boost achievement (Andrade et al., 2008; Kulkarni et al., 2013; Lawson et al., 2012), but such studies are sparse and few have been conducted in a middle school setting (Ross et al., 2002; Sadler & Good, 2006). This study also sought to fill a gap in the literature by adding to the knowledge of the effect of self- and peer assessment on middle school student achievement. This chapter presents an analysis of the quantitative pre- and posttest data collected before and after the implementation of the intervention.

Instrumentation

Data were collected twice, before and after the aforementioned 10 self- and peer-assessment training sessions, which were conducted with the treatment group over a 3-week period. Data measured three variables, students' mathematics achievement (POSTACH), ability to accurately self-assess (POSTSELF), and ability to accurately peer-assess (POSTPEER). Four covariates, previous mathematical achievement (NAPD), and pretest measures of all outcome variables (PREPEER, PRESELF, and PREACH) were also measured. Mathematics achievement was measured using the aforementioned assessment item developed by a team of mathematicians, mathematics educators, and middle school teachers, and the covariate, previous mathematical achievement, was measured by students' scores on the school-wide administration of the MAP test (NEA, 2012), which was administered in March, 2015. Students' abilities to accurately self- and peer-assess were both measured by the discrepancy between students' self- and peer-assessment scores and teacher scores, which were measured using the Self-Assessment Rubric and Teacher-/Peer-Assessment Rubric designed for this study.

Statistical Procedures

Rationale for Statistical Procedures

This study had an independent variable with two groups, treatment or no treatment, and three dependent variables, achievement, students' ability to accurately self-assess, and students' ability to accurately peer-assess. Covariates included previous mathematical achievement, and pre-test measures for each dependent variable. Given these factors, a multivariate analysis of covariance (MANCOVA) was deemed the appropriate way to analyze the data (Field, 2009; Hinton et al., 2014) and was conducted in SPSS (Field, 2009). In addition, the outcome variables were hierarchical, entailing a subsequent multivariate multilevel model (Baldwin, Imel, Braithwaite, & Atkins, 2014; Thum, 1997) to analyze the relationships among the outcome variables and covariates. When conducting all analysis, the standard α level of .05 and power of .8 were applied.

Assumptions

Assumptions related to multivariate analysis. The assumptions for multivariate analysis include independent observations, multivariate normality, and equal variance in the dependent variables across groups (Stevens, 2002). In this study, the covariate, independent, and dependent variable measures were independent and consisted of interval-level data. None of the data were normally distributed, according to the Shapiro-Wilk test, which was significant (Hinton et al., 2014). A visual inspection finds a reasonable level of normality in the post-achievement variable, although the other variable data were positively skewed. Figure 4.1 below reflects these results.

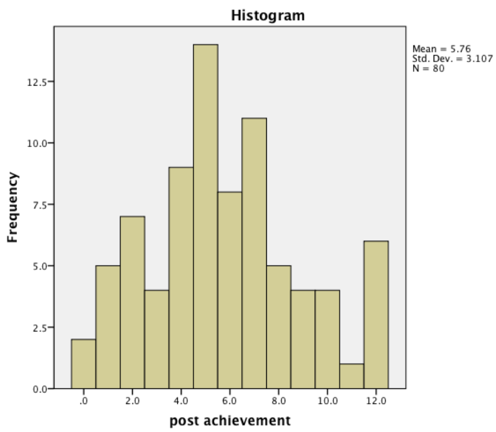


Figure 4.1 Distribution of post-achievement data.

As previously noted, student achievement and ability to assess have been positively linked in the research literature (Boud & Falchikov, 1989). Therefore, because the researcher purposively included the relatively few higher-achieving students available in order to test the treatment across achievement levels, non-normality was anticipated. Because there is little difference between actual and nominal α, multivariate tests are robust to violations of normality, with respect to Type 1 error, even with a small sample size (Stevens, 2002). To test the third assumption, the Levene's Tests of Equality of Error Variances was conducted on the dependent variables. The test was not significant for achievement and peer-assessment accuracy variable, but it was significant for the posttest measure of students' ability to accurately self-assess, indicating that the assumption was not met. However, multivariate analysis is robust to a violation of this assumption, as long as groups are equal or close to equal, with a ratio of the largest group to the smallest no larger than 1.5 (Stevens, 2002). In this case, the ratio of control group numbers to treatment group numbers is 41/39, approximately 1.08, which is within the allowable ratio.

Assumptions related to analysis of covariance. The aforementioned analysis of covariance was appropriate if a correlation did, in fact, exist between the potential covariates and the outcome variables (Stevens, 2002). Correlation coefficients (Muthén, L.K. & Muthén, 1998-2015) were calculated and are displayed in the table below. As the table reflects, the majority of the correlations were close to or greater than the benchmark for a medium-sized correlation, $r = .3$ (Cohen, 1992).

Table 4.1

Correlation Coefficients for Outcome Variables and Potential Covariates

Variable	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1. NAPD	1.000						
2. PREACH	0.548	1.000					
3. PRESELF	0.090	0.201	1.000				
4. PREPEER	0.017	0.169	-0.229	1.000			
5. POSTACH	0.647	0.789	0.219	0.133	1.000		
6. POSTSELF	0.239	0.527	0.351	0.069	0.642	1.000	
7. POSTPEER	0.086	-0.088	0.014	0.021	-0.047	-0.002	1.000

Findings

Descriptive Statistics

Mathematical achievement was measured by the aforementioned assessment and rubric, and the ability to accurately self- and peer-assess were measured by the discrepancy between the teacher and student rubric scores. Therefore, an increase in the achievement value reflects higher achievement, if a significant effect was found, while a decrease in the self-and peer-assessment discrepancy scores represents increased student ability to assess, if a significant effect was found.

Table 4.2

Descriptive Statistics for Variables by Treatment/Control Group

Pretest Measures				
	Treatment Group		**Control Group**	
	M	**SD**	**M**	**SD**
NAPD	1.74	0.75	1.66	0.73
PREACH	4.46	2.79	4.42	2.93
PRESELF	3.87	2.85	3.98	2.85
PREPEER	2.90	2.07	2.44	1.43
Posttest Measures				
	Treatment Group		**Control Group**	
	M	**SD**	**M**	**SD**
POSTACH	7.54	2.84	4.07	2.33

POSTSELF	2.23	2.00	4.05	2.91
POSTPEER	1.69	1.66	2.2	1.85

Main Findings

Wilk's Lambda (λ) indicated that there was some multivariate effect of the treatment on the dependent variables, $F(3,72) = 23.38$; $p < .001$, with a partial eta squared, $\eta^2 = .49$, that is just under the benchmark for a large effect size, .5 (Cohen, 1992).

Research Question 1

The first research question investigated the effect of the intervention of self- and peer-assessment training on middle school students' mathematics achievement. ANCOVA test data indicated a difference in students' mathematical achievement, based on their placement in the treatment group, with an *F* ratio $F(1,74) = 70.85$; $p < .001$, and a partial eta squared, $\eta^2 = .49$, that is just under the benchmark for a large effect size, .5 (Cohen, 1992).

Research Question 2

The second research question investigated the effect of the intervention of self- and peer-assessment training on middle school students' ability to accurately self-assess. The accuracy of students' self-assessment was measured by the discrepancy between the score they assigned to their own work and that assessed by the teacher. ANCOVA test data indicated a difference in students' ability to accurately self-assess, based on their placement in the treatment group, with an *F* ratio $F(1,74) = 12.46$; $p = .001$, and a partial eta squared, $\eta^2 = .14$, which met the benchmark for a small effect size, .1 (Cohen, 1992).

Research Question 3

The third research question investigated the effect of the intervention of self- and peer-assessment training on middle school students' ability to accurately peer-assess. The accuracy of students' self-assessment was measured by the discrepancy between the score they assigned to their peer's work and that assessed by the teacher. ANCOVA test data indicated no difference in students' ability to accurately peer-assess, based on their placement in the treatment group, with an *F* ratio $F(1,74) = 1.91$; $p = .171$, and a partial eta squared, $\eta^2 = .03$, which was lower than the benchmark for a small effect size.

Path Analysis of Multivariate Outcomes

Rationale. The intervention was developed to improve students' ability to self- and peer-assess (POSTSELF and POSTPEER) and, consequently, to elevate students' mathematics achievement (POSTACH). A reciprocal effect between self- and peer-assessment ability was also hypothesized. A path analysis of multivariate outcomes was used to investigate the strength of the various relationships that might exist among the variables (O'Rourke & Hatcher, 2013).

Recoding. Self- and peer-assessment discrepancy scores were recoded by subtracting each assessment score from 15, to allow a high self- or peer-assessment score to correlate with a high achievement score, producing a positive correlation. For example, an original POSTSELF score of 2 would translate to a recoded score of 15-2 = 13.

Best-fit model. A set of fit statistics was considered when determining the fit of the model. Although there is not a set benchmark for an acceptable chi-square test value, because it "assesses the magnitude of discrepancy between the sample and fitted covariance matrices" (Hooper, Coughlan, & Mullen, 2008, p. 53), the smaller it is the more accurate the model (Suhr, 2008). The chi-square for this model was 5.7. Because the comparative fit index (CFI) is a way of comparing, via ratio, the proposed model to a baseline model (Hooper, et al., 2008), the higher the value, the better the fit a model is for the data; the standard is a CFI value equal or greater than 0.90 (Suhr, 2008). The final model had a CFI of .99, which indicated a good fit. The Tucker Lewis index (TLI) is another statistic that indicates a better fit with a higher value, because it too is a method of comparing the model of interest to a more strict "null" model (Bollen, 1986); the recommended TLI is at or above .95 (Hooper, Coughlan, & Mullen, 2008). This model had a TLI of .95, indicating a good fit. Because it is a method of measuring the discrepancy between the correlation of the model and the correlation that as been predicted, the smaller a model's standardized root mean square residual (SRMR), the better its fit; the standard recommended value is .08 or lower (Kenny, 2014). This model's SRMR was .03, indicating a good fit. As an absolute measure of fit, which is based upon the notion that a best-fitting model has a value of 0, (Kenny, 2014), the smaller a model's root mean square error of approximation (RMSEA) the better the model fit; the standard is a value equal to or less than 0.06 (Suhr, 2008). This model had an RMSEA of .07 and a confidence interval ranging from .000 to .195. Although these RMSEA values are not ideal, they are acceptable in light of the other fit statistics and the relatively low sample size (Byrne, 2012), which likely negatively impacted them.

Path coefficients. Path coefficients, β, allowed the researcher to determine the specific impact one variable had on another, as well as the impact of the intervention on dependent variables (Wright, 1960); these are reflected here in Table 4.3. The model's significant unstandardized path coefficients are shown in Figure 4.2 below.

Table 4.3

Model Unstandardized Path Coefficients and p Values

Independent Variable	Dependent Variable	*β*	*p*
Treatment	POSTACH	2.702	.000
PREACH	POSTACH	.343	.000
POSTPEER	POSTACH	-.034	.735
POSTSELF	POSTACH	.371	.000
NAPD	POSTACH	1.066	.000
Treatment	POSTSELF	1.728	.000
PREACH	POSTSELF	.081	.456
PRESELF	POSTSELF	.360	.000
PREPEER	POSTSELF	.028	.837
NAPD	POSTSELF	.722	.054
Treatment	POSTPEER	.565	.138
PREACH	POSTPEER	-.128	.090
PRESELF	POSTPEER	.103	.179
PREPEER	POSTPEER	.146	.186

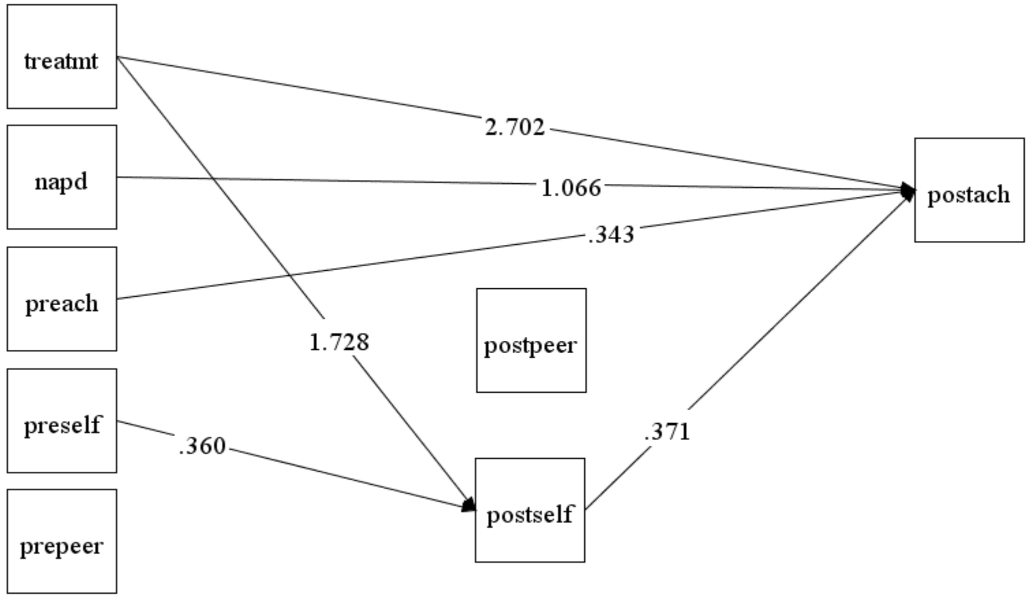


Figure 4.2. Significant path coefficients.

Because the analysis conducted to obtain these coefficients is a form of linear modeling, each significant *β* allows for the prediction of dependent variables, based on independent variable values (Baldwin et al., 2014; Thum, 1997). These unit changes represent a difference from the mean, as all independent variables are grand mean centered. For example, the coefficient (*β*) from the treatment to POSTACH is 2.702. The "average" student—that is, one who earned scores equal to the overall mean scores on each independent variable measure and who was in the treatment group—would be predicted to earn a POSTACH score equal to the overall mean for the treatment and control groups POSTACH score, 5.763, plus an additional 2.702 points, for a total POSTACH score of 8.465. When students do not earn the average overall score of certain independent variables, that deviation from the mean can be used to predict dependent variable scores, in that one unit change in the independent variable predicts a change in the dependent variable equal to the coefficient. Consider, for example, a student in the control group who might earn average scores on all pretest measures except for a PREACH score of 7, which is 2.563 points above the overall sample mean of 4.437. Because the coefficient from PREACH to POSTACH is .343, the student could be predicted to earn 2.563(.343) = .879 points above the group POSTACH mean, for a total score of 5.763 + .879 = 6.642.

Summary of Findings

Statistical tests revealed a positive effect of peer- and self-assessment training on students' mathematical achievement and ability to accurately self-assess but no effect on the accuracy of students' assessments of peer work. In addition, path analysis detected six significant interactions between variables. The following chapter will discuss this study's key findings and path analysis as they relate to previous research, implications for teacher practice, and potential future research.

CHAPTER V: DISCUSSION

Introduction

This chapter provides a summary of the study, results, and conclusions based on the findings presented in Chapter 4. Additionally, the chapter offers a discussion of the implications of these findings on teaching. Finally, the chapter concludes with recommendations for future research.

Summary of Study

Restatement of the Problem

Recent data on mathematics achievement in the United States (USED, 2013) indicate that U.S. students lag behind their international peers and thus may prove unprepared to successfully meet the demands of an increasingly global economy (NCTM, 2013a).

Restatement of Purpose and Research Questions

Previous research has demonstrated that the most significant opportunities for improved mathematics learning occur in the classroom (Black & Wiliam, 1998b; Noonan & Duncan, 2005), especially when teachers implement formative assessment strategies (Andrade & Valtcheva, 2009; Black & Wiliam, 1998b; Brown & Hirschfeld, 2007; Davis & McGowen, 2007; Heritage & Heritage, 2013; Leahy et al., 2005; Nicol & Macfarlane-Dick, 2006; Ross et al., 2002; Suurtamm et al., 2010; Thomas et al., 2011; Wiliam, 2007b). Also commonly accepted is the notion that making students full partners within a classroom formative assessment plan, through self- and peer assessment, results in positive outcomes for students (Heritage, 2007; Noonan & Duncan, 2005; Wiliam, 2007a).

Despite the advantages of students' self-assessment and peer assessment, teachers sometimes hesitate to implement these strategies, fearing that, generally, students are not equipped to accurately appraise their own work (Noonan & Duncan, 2005). The reluctance to implement student assessment strategies could be intensified among middle school teachers, by the fact that most of the studies that have examined teacher-student assessment agreement (Boud & Falchikov, 1989; Falchikov & Boud, 1989; Falchikov & Goldfinch, 2000), including the very few that have specifically investigated strategies that lessen the gap between teachers' and students' evaluations (Kulkarni et al., 2013; Lawson et al., 2012; Ross et al., 2002), have been confined to a higher-education setting. To fill this void in the self- and peer-assessment research, this study tested in a middle school setting a self- and peer-assessment-training instructional sequence that was a compilation of research-based strategies. The study sought to determine the effect of the instructional sequence both on middle school students' ability to accurately self-assess and peer-assess and on their mathematics achievement.

Review of Methodology

Type of research. This was a quantitative study, with the outcome variables students' ability to accurately self-assess, students' ability to accurately peer-assess, and students' mathematics achievement, as well as four covariates, previous mathematics achievement and pretest measures of the outcome variables; all were assessed numerically and analyzed using a multivariate statistical procedure (Cresswell, 2009; Teddlie & Tashakkori, 2009).

Sample and instrumentation. Data were collected twice, from 39 treatment group students and 41 control group students, who made up the convenience- and purposive-selected and assigned sample from a population of 7th-grade students at an urban middle school in a metropolitan area in the Midwest. These data were collected before and after the aforementioned 10 self- and peer-assessment-training sessions, which were conducted with the treatment group over a 3-week period. Mathematics achievement was measured using an assessment item developed by a team of mathematicians, mathematics educators, and middle school teachers, and the covariate, previous mathematical achievement, was measured by students' scores on the most recent school-wide administration of the MAP test (NEA, 2012). Students' abilities to accurately self- and peer-assess were measured by the discrepancy between students' self- and peer-assessment scores and teacher scores; all scores were calculated using rubrics designed by a National Board-certified middle school mathematics teacher with the input of three other middle school mathematics teachers and one mathematics teacher educator.

Statistical analysis. Because the study involved an independent variable with two groups, treatment and control, three outcome variables, and four covariates, a multivariate analysis of covariance (MANCOVA) was deemed the appropriate way to analyze the data (Field, 2009; Hinton et al., 2014). Moreover, because the outcome variables were hierarchical, a multivariate, multilevel model was necessary for the analysis (Baldwin et al., 2014; Thum, 1997). When conducting all analyses, the standards of an α level of .05 and a power level of .8 were applied.

Summary of Findings

Statistical tests revealed a positive effect of peer- and self-assessment training on students' mathematical achievement and on students' ability to accurately self-assess. The training did not, however, impact the accuracy of students' peer assessments. A path analysis was also conducted to determine the strength and direction of multiple combinations of variables; the analysis produced six significant path coefficients. A discussion of this study's findings as they relate to the previous research follows.

Impact of Self- and Peer-assessment Training on Mathematics Achievement

The intervention's positive impact on students' mathematics achievement and the medium to large effect size, partial $\eta^2 = .49$, are consistent with past studies, which have shown a link between student engagement in the assessment process and student achievement (Heritage, 2007; Noonan & Duncan, 2005; Wiliam, 2007a).

Intervention components. This study's intervention design included a combination of the following five strategies:

- defining the purpose and communicating the learning goals of the intervention (Brophy et al., 1983; Clark, 2012);
- communicating criteria for success (Black & Wiliam, 2009; Ross, 2006; Thomas et al., 2011; Wiliam, 2007a) and enlisting students in the development of criteria (Andrade & Du, 2007; Heritage, 2007; Nicol & Macfarlane-Dick, 2006; Ross et al., 2002);
- modeling for students how to apply assessment criteria (Ross, 2006; Ross et al., 2002) and offering them multiple opportunities to practice applying said criteria (Ross, 2006);

- providing students ample teacher and peer feedback (Andrade & Valtcheva, 2009; Black & Wiliam, 1998b; Clark, 2012; Davis & McGowen, 2007; Leahy et al., 2005; Nicol & Macfarlane-Dick, 2006; Ross et al., 2002; Sadler, 1989; Stiggins, 2009; Wiliam, 2007a); and
- allowing students opportunities to revise and resubmit work (Boud, 2000; Clark, 2012; Guskey, 2003; Nicol & Macfarlane-Dick, 2006; Noonan & Duncan, 2005; Ross, 2006; Sadler, 1989; Stiggins, 2002; Zimmerman, 2001).

Possible effects of intervention components on achievement. Although this quantitative study did not provided qualitative data to support the following hypotheses, the literature advances specific potential sources of this intervention's impact on achievement.

Defining the purpose and communicating the learning goals of the intervention. Because students often doubt the value of novel experiences (Black & Wiliam, 1998b), teachers must combat their resistance by clarifying for students lasting relevance and applicability of such experiences (Brophy et al., 1983). To the researcher's knowledge, the students in this study had very little, if any, previous or consistent experience with student assessment. The previous literature supports the possibility that some of the positive impact of the intervention could have resulted from students' enhanced buy-in to the relevance of assessment; this buy-in has been shown to promote heightened levels of engagement (Ross, 2006) and positive feelings toward assessment (Black & Wiliam, 1998a).

Communicating criteria for success and enlisting students in the development of criteria. Rubric use gave students successful-task-completion indicators and guidance that they might not have developed on their own, due to common misunderstandings of, or lack of exposure to, the qualities of good work (Ross, 2006). In addition, student cooperation in developing goals and identifying specific attributes of exemplary task completion was solicited and obtained, to instill them with a more thorough understanding of goals and criteria, another factor positively correlated with achievement (Andrade & Du, 2007; Black & Wiliam, 2009; Heritage, 2007; Nicol & Macfarlane-Dick, 2006; Ross, 2006; Ross et al., 2002; Thomas et al., 2011; Wiliam, 2007a). The historically low achievement of the majority of the sample students before the intervention, as measured by their MAP scores, could indicate a lack of understanding of the features of good work. Therefore, gains in student achievement might have been attained by exposing students to high-quality work and by explicating criteria. An added benefit of encouraging students to develop criteria within this study's large-group setting was the student-to-student and student-to-teacher dialogue; as students negotiate with peers and teachers, they actively construct knowledge and frequently experience enhanced levels of metacognition, two more outcomes tied to achievement (Black & Wiliam, 2009).

Modeling for students how to apply assessment criteria and offering them multiple opportunities to practice applying said criteria. By modeling for students how to apply assessment criteria (Ross, 2006; Ross et al., 2002) and giving them opportunities to practice the skill (Ross, 2006), teachers cultivate a better understanding of appropriate performance standards, laying the groundwork for subsequent student achievement.

Conversely, students who struggle tend to have a particularly limited understanding of criteria (Leahy et al., 2005). The researcher's use of these strategies, therefore, could have extended the student groups' understanding and thus contributed to their subsequent achievement.

Providing students ample teacher and peer feedback. Previous research has shown the value of applying criteria to peer work, citing that it allows students to exercise and sharpen their evaluation skills without the negative emotions that can sometimes come when analyzing one's own product (Leahy et al., 2005; Wiliam, 2007a). This benefit of peer work likely helped the, overall, low-achieving students in this study, because their negative past experiences with and resulting feelings about assessment could have impaired their willingness to assess their own work. Moreover, the whole-group setting, with dialogue directed towards assessment criteria and student work rather than alleged qualities of or behaviors exhibited by students themselves, provided a safe environment for students to engage with each other and the teacher; as indicated throughout, such engagement has been found to induce meaningful and frequent reflection (Black & Wiliam, 2009). While assessing work according to criteria within a group setting, students received constant teacher and peer feedback that allowed them to gauge and reflect upon the quality of their assessments, with an eye toward improvement.

Allowing students opportunities to revise and resubmit work. The intervention afforded students the opportunity to revise and resubmit their work after teacher instruction and, thereby, likely motivated them to engage in the process of attempting to improve their products, with greater buy-in of and a mastery-oriented focus toward assessments (Boud, 2000; Clark, 2012; Guskey, 2003; Nicol & Macfarlane-Dick, 2006;

Noonan & Duncan, 2005; Ross, 2006; Sadler, 1999; Stiggins, 2002; Zimmerman, 2001). Revision and resubmission, and the mastery focus on assessment (Stiggins, 2009), were particularly essential for this group of students, whose previous negative assessment experiences could have affected their motivation.

Implications of Study Intervention Results for Educators

Impact of self- and peer-assessment training on students' achievement. The impact of this intervention on achievement indicates that teachers should implement student assessment strategies in regular classroom practice. The aforementioned path analysis showed that the treatment could be predicted to directly contribute, for the average student, an achievement increase equal to almost 50% of the overall mean post-achievement score. Because many pre-service programs and much current professional development have left teachers untrained in effective student assessment practices, assessment strategies are often unused (Black & Wiliam, 1998b; Stiggins, 2002). For untrained teachers, this study contributes an explicit and demonstrably effective instructional sequence and exemplifies successful strategy implementation. Moreover, because student achievement gains were seen after just 10 hour-long sessions, the short duration of the intervention, arguably a limitation of the study, might move teachers once discouraged by school-day time constraints to pilot student assessment strategies. Because attempting too many new strategies at once could overwhelm teachers and students alike (Black & Wiliam, 1998b), one variation to the intervention might be to introduce assessment practices piecemeal (Leahy et al., 2005).

Impact of self- and peer-assessment training on students' assessment accuracy. In addition to investigating the intervention's impact on achievement, another objective of the study was to investigate the intervention's potential to close the gaps between student and teacher assessments. Previous attempts to more closely align peer assessments (Kulkarni et al., 2013) and self-assessments (Lawson et al., 2012) were successful. The outcome of this study's intervention was consistent with previous findings that students' assessments can become more closely aligned with those of their teachers (Kulkarni et al., 2013; Lawson et al., 2012; Sadler & Good, 2006); however, the study might fail to assuage all teachers' concerns about students' assessment accuracy, because the intervention did not close the gap between teacher and students' assessments of peer work. Despite some potential lingering teacher doubts about student assessment accuracy, gains in achievement obtained by the instructional sequence indicate a strong enough outcome for teachers to implement assessment training in classrooms on a more regular basis. Support for this argument may be traced back to Heritage (2007):

> "[E]ven though formative assessment strategies will not always meet accepted standards of validity and reliability, teachers need to understand that the quality of the assessment is an important concern. The overriding issue is consequential validity. Because the purpose of formative assessment is to promote further learning, its validity hinges on how effectively learning takes place in subsequent instruction." (p. 143)

Implications of intervention results with study sample. Previous research that links students' past achievement to their subsequent achievement and assessment accuracy (Boud & Falchikov, 1989; Falchikov & Boud, 1989) could cause some teachers to suspect that historically low-achieving students are unlikely to realize gains from participation in assessment. But the treatment students' development of enhanced assessment accuracy, despite the fact that 32 of the 39 had not previously demonstrated proficient levels of mathematics achievement, might persuade teachers that students of all abilities can hone their evaluative skills. In addition, although path coefficients found significant impacts of this study students' previous achievement on their post-achievement, the impact of the intervention on post-achievement (2.702) was almost three times that of students' MAP scores (1.066) and almost eight times that of their pre-assessment scores (.343). In other words, including students in the assessment process can potentially help them overcome barriers to future high achievement irrespective of their past performance.

Study Results and the Original Study Conjecture

Unexpected peer assessment results. As previously noted, the intervention did not affect students' ability to accurately peer-assess, and path analysis showed no impacts between students' self- and peer-assessment ability. This contradicted the researcher's foundational conjecture that self- and peer-assessment ability would rise commensurately and reciprocally. The result is counterintuitive, since increased ability to self-assess, it may be hypothesized, is indicative of students' heightened ability to evaluate the quality of student work in general, including that of peers. The path analysis results—which showed no link between pre-test peer assessment ability and posttest peer assessment ability, despite the significant links that can be seen between pre- and posttest achievement measures and pre- and posttest self-assessment measures--are likewise difficult to account for.

One possible explanation for the lack of an increase in peer-assessment accuracy, as well as the lack of a link between pre- and posttest peer assessment measures, could be that the peer assessment results did not accurately capture students' ability to assess peer work. Reflection on the intervention reveals a few causes for this possibility. Underlying these explanations is Stiggins's (2009) observation that students' beliefs about assessment can significantly impact their performance thereon; the manner in which peer assessment was introduced to students could have impeded student buy-in to the peer-assessment process relative to that of other components of the intervention, potentially influencing the peer-assessment results. When setting the purpose in the opening session of the intervention, much emphasis was placed on the value of self-assessment, but less on that of peer assessment. Possibly, the researcher did not provide convincing evidence to the students that peer assessment would have meaning for them. More time might have been spent conveying to students the fact that their achievement can be enhanced through assessment, even when they are not assessing their own work. The researcher could have also built into the instructional sequence more accountability for peer assessments, based on increased levels of accuracy that should occur across the duration of the intervention. Perhaps the intervention length needed to be longer, in order to allow students more time to become accustomed to and appreciative of peer assessment. Another potential factor could have been the manner in which data were collected. As previously mentioned, students completed an assessment task, then self-assessed their task, and, finally, peer assessed a simulated peer's work. The fact that the peer assessment was the last of three assessment tasks could have contributed to a lack of student motivation to complete it to the best of their ability.

The relatively small sample size of this study and the resulting low power of the peer assessment portion of the statistical analysis could also have contributed to the lack of a positive impact on peer assessment accuracy. The peer assessment ANCOVA analysis produced observed power of .28, much lower than the standard .8 (Field, 2009). In other words, if the intervention did, in fact, have an effect on peer assessment accuracy, there was only a 28% chance, with this particular sample, that the effect would have been detected.

Expected self-assessment results. Although the anticipated effects of the intervention on peer assessment accuracy was not found, the study did identify several effects related to self-assessment accuracy. The path analysis found an impact of students' PRESELF on their POSTSELF, which was expected. The treatment also had an effect on POSTSELF, illustrated by a coefficient, 1.728, that is almost five times the coefficient linking PRESELF and POSTSELF, .360. This result is consistent with the previously described treatment impact on POSTACH, which, similarly, indicates the intervention's potential to help students overcome previous assessment deficits. The direct impact of the intervention on POSTACH was supplemented by an indirect one, as well, since the path from treatment to POSTSELF is continued by a path from POSTSELF to POSTACH, with a coefficient .371. In other words, a student's participation in the intervention resulted in higher predicted POSTSELF, which led to higher predicted POSTACH. This finding was consistent with the original study conjecture.

Recommendations for Future Research

Reflection on and acknowledgement of the study limitations, as well as indications of some positive impact despite its limitations, yield recommendations for future scholarly research.

Inclusion of Qualitative Data

Although this study's quantitative data revealed important insights into the potential power of student assessment techniques, they simultaneously open the door to many other questions, many of which might have been answered by introducing qualitative measures. Without these, the possible effects of intervention components on achievement and student traits, discussed above, are no more than that: "possible effects." For example, with no recorded evidence of students' revised work subsequent to receiving teacher feedback, the impact of the feedback can be speculated but not definitively claimed. Similarly, even if the feedback did, in fact, have an effect grounded in qualitative evidence, the lack of qualitative evidence of the specific characteristics of the feedback, again, raises speculation but denies resolution. Interviews with students, observations of the school and classroom environments, or recordings of intervention sessions could have provided evidence of the heightened levels of student engagement and buy-in to the assessment process that are hypothesized to exist based on the assessment results. Similar qualitative data could have also identified environmental, student, or intervention-implementation factors that may explain why students more effectively evaluated their own work than that of their peers.

Examination of Individual Components of the Intervention

Because the intervention was a compilation of the five formative assessment strategies, and because these strategies did not have independent measures, the study could not isolate whether any single strategy had a greater impact on the outcome. To better differentiate the strategies, future research might consider potential effects on students' achievement and assessment ability at each individual stage in the intervention sequence. Examining the strategies in isolation would also make more feasible the much-needed inclusion of qualitative data in future studies.

Timing and Duration of Intervention Implementation

One could consider examining the effects of the intervention after it is implemented at the beginning of a school year and revisited throughout. Teachers could solidify students' belief in the value of the assessment process, and secure their active participation therein, by introducing the intervention when it has the potential to impact their achievement over the course of the entire school year, namely, at the beginning of the term. By implementing the intervention at the beginning of the school year and revisiting it throughout, the components of the intervention would be tested over a longer period of time, thus potentially enhancing the positive effects this intervention was able to generate in a short duration.

Data Collection Procedure Changes

As previously noted, the manner in which the data were collected, especially the fact that peer assessment data were collected last during the pre- and posttest administrations, could have affected the accuracy of the data. In order to help ensure that data results are not the result of extraneous factors, such as timing of assessments, future researchers should not collect self-assessment and peer assessment from all students in the same order.

Role of Student Demographic or Personal Factors

Previous mathematical achievement was the only individual student factor examined in this study; future research might consider a greater range of factors. For example, student characteristics such as race, gender, and socioeconomic status could relate to the effects of various intervention strategies. In addition, potentially influential student factors such as motivation and attitudes toward mathematics could be observed and tracked throughout the course of an assessment training sequence.

Greater Range of Student Ability Levels

The achievement levels of the students at the study site were low overall; therefore, the intervention could not be tested across a wide range of previous achievement levels. Future research should test the intervention with a more academically diverse sample, as to aid in determining whether strategies within the intervention can close achievement gaps, as previous research suggests (Black & Wiliam, 1998a), or even expand the capacity of already high-achieving students.

Greater Power

The low sample size and consequent low power of the intervention could explain the absence of a peer-assessment effect. The power analysis was also based on research that was, arguably, not germane to a middle school setting. A higher sample size, therefore, could produce the significant effect on peer assessment that this study did not detect.

Conclusion

Despite seeming to have created more questions that it has answered, this study has, ultimately, contributed evidence to the existing literature that students' participation in the assessment process yields positive outcomes, most significantly, increased student achievement. The study has also added evidence that students who can more effectively self-assess also succeed more academically. Although there is still much investigation needed into specific aspects of student assessment practices, the implications from this study for teacher practice are clear: To promote the best achievement results for their students, teachers must implement strategies that make students full partners in classroom assessment.

REFERENCES

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*(3), 261-271.

Andrade, H., & Boulay, B. (2003). Role of rubric-referenced self-assessment in learning to write. *Journal of Educational Research, 97*(1), 21-34.

Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment and Evaluation in Higher Education, 32*(2), 159-181.

Andrade, H., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-centered self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practices, 27*(2), 3-13.

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Into Practice, 48,* 12-19. doi: 10.1080/00405840802577544

Baartman, L., & Ruijs, L. (2011). Comparing students' perceived and actual competence in higher vocational education. *Assessment and Evaluation in Higher Education, 36*(4), 385-398.

Baldwin, S. A., Imel, Z. E., Braithwaite, S. R., & Atkins, D. C. (2014). Analyzing multiple outcomes in clinical research using multivariate multilevel models. *Journal of Consulting and Clinical Psychology, 82*(5), 920-930. doi:10.1037/a0035628.

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes, 50,* 248-287.

Bandura, A. (2006). Adolescent development from an agentic perspective. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 1–43). Greenwich, CT: Information Age.

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61,* 213-238.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy, & Practice, 5*(1), 7-74.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139-148.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation, and Accountability, 1*(1), 1-40. doi: 10.1007/s11092-008-9068-5

Blumenfeld, P., Soloway, E., Marx, R., Krajcik, J., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist, 26*(3 & 4), 369-398.

Bollen, K. A. (1986). Sample size and Bentler and Bonett's nonnormed fit index. *Psychometrika, 51*(3), 375-377. Retrieved from http://link.springer.com/article/10.1007/BF02294061#page-1

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education, 22*(2), 151-167.

Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education, 18*(5), 529-549.

Brophy, J., Rohrkemper, M., Rashid, H., & Goldberger, M. (1983). Relationships between teachers' presentations of classroom tasks and students' engagement in those tasks. *Journal of Educational Psychology, 75*(4), 544-552.

Brown, G. T. L., & Hirschfeld, G. H. F. (2007). Students' conception of assessment and mathematics: Self-regulation raises achievement. *Australian Journal of Educational Development Psychology, 7,* 63-74.

Bryant, D. A., & Carless, D. R. (2010). Peer assessment in a test-dominated setting: Empowering, boring, or facilitating examination preparation?. *Educational Research for Policy and Practice, 9*(1), 3-15.

Butler, R. (1990). The effects of mastery and competitive conditions on self-assessment at different ages. *Child Development, 61,* 201-210.

Butler, D., & Winne, P. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245-281.

Byrne, B. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York, NY: Routledge.

Cantrell, S. C., Burns, L. D., & Callaway, P. (2009). Middle- and high-school content area teachers' perceptions about literacy teaching and learning. *Literacy Research and Instruction, 48*(1), 76-94.

Chatterji, M. (2003). *Designing and using tools for educational assessment.* Boston, MA: Allyn.

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

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review, 24,* 205-249.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Conley, D. (2011). Building on the Common Core. *Educational Leadership, 68*(6), 16-20.

Credé, M., & Kuncel, N. (2008). Study habits, skills, and attitudes: The third pillar supporting collegiate academic performance. *Perspectives on Psychological Science, 3*(6), 425-453.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches.* Los Angeles, CA: Sage.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*(4), 438-481.

Curtis, D. A., Lind, S. L., Dellinges, M., Setia, G., & Finzen, F. (2008). Dental students' self-assessment of preclinical examinations. *Journal of Dental Education, 72*(3), 265-277.

Davis, G. E., & McGowen, M. A. (2007). Formative feedback and the mindful teaching of mathematics. *Australian Senior Mathematics Journal, 21*(1), 19-30.

Deakin-Crick, R., Sebba, J., Harlen, W., Yu, G., & Lawson, H. (2005). *Systematic review of research evidence of the impact on students of self- and peer assessment.* London, UK: EPPI Centre, Social Science Research Unit, Institute of Education, University of London. Retrieved from http://www.eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=HKKPQEO72Ss%3D&tabid=2415&mid=4477

DiPrete, T. A., & Gangl, M. (2004). Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. *Sociological Methodology, 34*(1), 271-310.

Ellis, C., & Folley, S. (2010). Using student assessment choice and eAssessment to achieve self-regulated learning. In G. Dettori & D. Persico (Eds.), *Fostering self-regulated learning through ICT* (pp. 89-104). New York, NY: IGI.

Eva, K. W., & Regehr, G. (2011). Exploring the divergence between self-assessment and self-monitoring. *Advances in Health Science Education, 16,* 311-329. doi: 10.1007/s10459-010-9263-2

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59*(4), 395-430.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*(3), 287-322.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London, UK: Sage.

Fox, S., & Dinur, Y. (1988). Validity of self-assessment: A field evaluation. *Personnel Psychology, 41*(3), 581-592.

Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education, 1*(1), 3-31.

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction, 20,* 304-315.

Guskey, T. (2003). How classroom assessments improve learning. *Educational Leadership, 60*(5), 6-11.

Halani, A., Davis, O., & Roh, K. H. (2013). Critiquing the reasoning of others: Devil's advocate and peer interpretations as instructional interventions. Proceedings of the 16th Annual Conference on Research in Undergraduate Mathematics Education. Retrieved from https://scholar.google.com/scholar?hl=en&as_sdt=0,18&q=CRITIQUING+THE+REASONING+OF+OTHERS%3A+DEVIL'S+ADVOCATE+AND+PEER+INTERPRETATIONS+AS+INSTRUCTIONAL+INTERVENTIONS

Hall, T. (2002). *Differentiated instruction.* Effective classroom practices report. National Center on Accessing the General Curriculum, CAST, U.S. Office of Special Education Programs.

Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education, 4*(3), 365-379.

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

Heritage, M. (2010a). *Formative assessment: Making it happen in the classroom*. Thousand Oaks, CA: Corwin.

Heritage, M. (2010b). Formative assessment and next-generation assessment systems: Are we losing an opportunity? National Center for Research on Evaluation, Standards, and Student Testing (CRESST) and the Council of Chief State School Officers (CCSSO). Washington, DC: CCSSO. Retrieved from http://eric.ed.gov/?id=ED543063

Heritage, M., & Heritage, J. (2013). Teacher questioning: The epicenter of instruction and assessment. *Applied Measurement in Education, 26*(3), 176-190. doi: 10.1080/08957347.2013.793190

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371-404). Greenwich, CT: Information Age.

Hinton, P. R., McMurray, I., & Brownlow, C. (2014). SPSS explained (2nd ed.). New York, NY: Routledge.

Hooper, D., Coughlan, J. and Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods, 6*(1), 53-60.

Hwang, W., Chen, N., Shadiev, R., & Li, J. (2011). Effects of reviewing annotations and homework solutions on math achievement. *British Journal of Educational Technology*, *42*(6), 1016-1028.

Johnson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity, and educational consequences. *Educational Research Review, 2,* 130-144.

Kenny, D. A. (2014). *Measuring Model Fit*. Retrieved from http://davidakenny.net/cm/fit.htm

Kentucky Department of Education (2014a). *Kentucky school report card.* Retrieved from https://applications.education.ky.gov/src/Profile.aspx

Kentucky Department of Education (2014b). *School report card archive 2006-2011.* Retrieved from https://applications.education.ky.gov/schoolreportcardarchive/

Kentucky Department of Education (2014c). *K-Prep.* Retrieved from http://education.ky.gov/AA/Assessments/Pages/K-PREP.aspx

Kim, C. & Keller, J. M. (2008). Effects of motivational and volitional email messages (MVEM) with personal messages on undergraduate students' motivation, study habits and achievement. *British Journal of Educational Technology*, 39(1), 36-51.

Kimberlin, C. L. & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health System Pharmacists, 65,* 2276-2284. Retrieved from http://www.ajhepworth.yolasite.com/resources/9817-Reliabillity%20and%20validity.pdf

Kirby, N. F., & Downs, C. T. (2007). Self-assessment and the disadvantaged student: Potential for encouraging self-regulated learning?. *Assessment and Evaluation in Higher Education, 32*(4), 475-494.

Kirschenbaum, D., & Karoly, P. (1977). When self-regulation fails: Tests of some preliminary hypotheses. *Journal of Consulting and Clinical Psychology, 45*(6), 1116-1125.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, meta-analysis, and a preliminary feedback intervention theory. *Psychological Review, 119*(2), 254-284.

Koth, C. W., Bradshaw, C. P., & Leaf, P. J. (2008). A multilevel study of predictors of student perceptions of school climate: The effect of classroom-level factors. *Journal of Educational Psychology, 100*(1), 96-104.

Kulkarni, C., Wei, K. P., Le, H., Chia, D., Papadopoulos, K., Cheng, J., Koller, D., & Klemmer, S. R. (2013). Peer and self assessment in massive online classes. *ACM Transactions on Computer-Human Interaction, 9*(4). Retrieved from http://hci.stanford.edu/publications/2013/Kulkarni-peerassessment.pdf

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity, and educational consequences. *Educational Research Review, 2,* 130-144.

Lawson, R. J., Taylor, T. L., Thompson, D. G., Simpson, L., Freeman, M., Treleaven, L., & Rohde, F. (2012). Engaging with graduate attributes through encouraging accurate self-assessment. *Asian Social Science, 8*(4), 3-12.

Leach, L. (2012). Optional self-assessment: Some tensions and dilemmas. *Assessment and Evaluation in Higher Education, 37*(2), 137-147.

Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership, 63*(3), 18-24.

Leahy, S., & Wiliam, D. (2012). Scaling up professional development for formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 49-71). Thousand Oaks, CA: Sage.

Liu, M. F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279-290.

MacDonald, K. (2011). A reflection on the introduction of a peer and self assessment initiative. *Practice and Evidence of Scholarship of Teaching and Learning in Higher Education, 6*(1), 27-42.

Mathematics Common Core Coalition. (n.d.). Frequently asked questions about the Common Core Math Standards. Retrieved from http://www.nctm.org/mathcommoncore/

McGregor, T. (2007). *Comprehension connections: Bridges to strategic reading*. Portsmouth, NH: Heinemann.

McMillan, J. (2001a). *Classroom assessment: Principles and practice for effective instruction* (2nd ed.). Boston, MA: Allyn.

McMillan, J. (2001b). *Essential assessment concepts for teachers and administrators*. Thousand Oaks, CA: Corwin.

Mertler, C., & Vannatta, R. (2002). *Statistical methods: Practical application and interpretation* (2nd ed.). Los Angeles, CA: Pyrczak.

Moskal, B. M. (2000). Scoring rubrics: What, when, and how?. *Practical Assessment, Research & Evaluation, 7*(3). Retrieved from http://PAREonline.net/getvn.asp?v=7&n=3.

Moskal, B. M. (2003). Recommendations for developing classroom performance assessments and scoring rubrics. *Practical Assessment, Research & Evaluation, 8*(14). Retrieved from http://PAREonline.net/getvn.asp?v=8 &n=14.

Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation, 7*(10). Retrieved from http://PAREonline.net/getvn.asp?v=7&n=10 .

Muthén, B. O. (1990). *Multilevel factor analysis of class and student achievement components* (Report No. 332). Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing. Retrieved from http://files.eric.ed.gov/fulltext/ED341735.pdf

Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Authors.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. V. Merrienboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed.) (pp. 125-143). New York, NY: Macmillan.

National Council of Supervisors of Mathematics & Association of Mathematics Teacher Educators. (2014). *Position: Improving student achievement in mathematics through formative assessment in instruction.* Retrieved from http://amte.net/publications/position-statements

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2013a). *Position statement on formative assessment.* Retrieved from http://www.nctm.org/Standards-and-Positions/Position-Statements/Formative-Assessment/

National Council of Teachers of Mathematics. (2013b). *Position statement on supporting the Common Core state standards for mathematics.* Retrieved from http://www.nctm.org/Standards-and-Positions/Position-Statements/Supporting-the-Common-Core-State-Standards-for-Mathematics/

National Council of Teachers of Mathematics. (2013c). *Stay the course: Supporting success with the Common Core State Standards*. Retrieved from http://www.nctm.org/News-and-Calendar/Messages-from-the-President/Archive/Linda-M_-Gojak/Stay-the-Course_-Supporting-Success-with-the-Common-Core-State-Standards/

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core state standards for mathematics.* Washington, DC: Authors.

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. J. Kilpatrick, J. Swafford, and B. Findell (Eds.). Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy.

National Science and Technology Council. (2013). *Federal science, technology, engineering, and mathematics (STEM) education 5-year strategic plan.* Retrieved from http://www.whitehouse.gov/sites/default/files/microsites/ostp/stem_stratplan_2013.pdf

Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31,* 199-218.

Noonan, B., & Duncan, C. R. (2005). Peer and self-assessment in high schools. *Practical Assessment, Research & Evaluation, 10*(17). Retrieved from http://pareonline.net/pdf/v10n17.pdf

Northwest Evaluation Association. (2012). *Teacher hand-book: Measures of academic progress (MAP)*. Portland, OR: Author. Retrieved from http://legacysupport.nwea.org/sites/www.nwea.org/files/resources/Teacher%20Handbook_0.pdf

Northwest Evaluation Association. (2013). *Kentucky linking study*. Portland, OR: Author. Retrieved from https://www.nwea.org/content/uploads/2013/03/KY_2012_LinkingStudy.pdf

Nulty, D. (2011). Peer and self-assessment in the first year of university. *Assessment and Evaluation in Higher Education, 36*(5), 493-507.

Onion, A., & Javaheri, E. (2011). Self and peer assessment of mathematical processes. *Mathematics Teaching, 105*(2), 30-32.

O'Rourke, N., & Hatcher, L. (2013). *A step-by-step approach to using SAS for factor analysis and structural equation modeling* (2nd ed.). Cary, North Carolina: SAS Institute Inc.

Orsmond, P., Merry, S., & Reiling, K. (2002). The use of exemplars and formative feedback when using student derived marking criteria in peer and self-assessment. *Assessment in Higher Education, 27*(4), 309-323.

Pakaslahti, L., & Keltikangas-Järvinen, L. (2000). Comparison of peer, teacher and self-assessments on adolescent direct and indirect aggression. *Educational Psychology, 20*(2), 177-190.

Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice, 28*(3), 5-13.

Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007). *The role of interim assessments in a comprehensive assessment system: A policy brief.* Retrieved from: http://www.nciea.org/publications/PolicyBriefFINAL.pdf

Perry, N. E., VandeKamp, K. O., Mercer, L. K., & Nordby, C. J. (2010). Investigating teacher-student interactions that foster self-regulated learning. *Educational Psychologist, 37*(1), 5-15.

Peters, M. (1999). Poststructuralism and education. In M. Peters, T. Besley, A. Gibbons, B. Žarnić, & P. Ghiraldelli (Eds.), *The Encyclopaedia of Educational Philosophy and Theory*. Retrieved from http://eepat.net/doku.php?id=poststructuralism_and_philosophy_of_education.

Phielix, C., Prins, F., Kirschner, P., Erkens, G., & Jaspers, J. (2010). Group awareness of social and cognitive performance in a CSL environment: Effects of peer feedback and reflection tool. *Computers in Human Behavior, 27*(3), 1087-1102.

Pintrich, P. R., & Zusho, A. (2007). Student motivation and self-regulated learning in the college classroom. In P. Raymond & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 731-810). Dordrecht, The Netherlands: Springer.

Popham, W. J. (2011). Formative assessment—A process, not a test. *Education Week, 30*(21), 35-37.

Porter, A., McMaken, J., Hwang, J., & Yang, R. (2011). *Educational Researcher, 40*(3), 103-116.

Porter-O'Donnell, C. (2004). Beyond the yellow highlighter: Teaching annotation skills to improve reading comprehension. *English Journal, 93*(5), 82-89.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science, 28*(1), 4-13.

Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., & Kuts, C. (2002). A neural basis for social operation. *Neuron, 35,* 395-405.

Ross, J., Rolheiser, C., and Hogaboam-Gray, A. (1999). Effects of self-evaluation training on narrative writing. *Assessment Writing, 6*(1), 107-132.

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research & Evaluation, 11*(10). Retrieved from https://tspace.library.utoronto.ca/bitstream/1807/30005/1/The%20Reliability,%20Validity,%20and%20Utility%20of%20Self-Assessment.pdf

Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002). Student self-evaluation in grade 5-6 mathematics effects on problem-solving achievement. *Educational Assessment, 8*(1), 43-59.

Rust, C., Price, M., & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment and Evaluation in Higher Education, 28*(2), 147-164.

Ryan, A., Hicks, L., & Midgley, C. (1997). Social goals, academic goal, and avoiding seeking help in the classroom. *Journal of Early Adolescence, 17*(2), 152-171.

Ryan, A. M., & Patrick, H. (2001). The classroom social environment and changes in adolescents' motivation and engagement during middle school. *American Educational Research Journal, 38*(2), 437-460.

Saderholm, J., Ronau, R., Brown, E. T., & Collins, G. (2010). Validation of the Diagnostic Teacher Assessment of Mathematics and Science (DTAMS) instrument. *School Science and Mathematics, 110*(4), 180-192.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18,* 119-144.

Sadler, P., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment, 11*(1), 1-31.

Schunk, D. (2012). *Learning theories: An educational perspective* (6th ed.). Boston, MA: Pearson.

Schoenfeld, A. H. (2015). Summative and formative assessments in mathematics supporting the goals of the Common Core standards. *Theory Into Practice, 54*(3), 183-194.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton.

Shepard, L. A. (2009). Commentary: Evaluating the validity of formative and interim assessment. *Educational Measurement: Issues and Practice, 28*(3), 32-37.

Sluijsmans, D., Dochy, F., & Moerkerke, G. (1999). Creating a learning environment by using self-, peer- and co-assessment. *Learning Environments Research, 1,* 293-319.

Statistics Solutions (2014). Sample size write-up and generator and resources. Retrieved from http://www.statisticssolutions.com/mancova-2-levels-and-3-dependent-variables/

Stepanek, J., & Jarrett, D. (1997). *Assessment strategies to inform science and mathematics instruction: It's just good teaching*. Portland, OR: Northwest Regional Education Laboratory.

Stevens, J. (2002). *Applied multivariate statistics for the social sciences*. Mahwah, N.J: Erlbaum.

Stiggins, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan, 83*(10), 758-765.

Stiggins, R. (2009). Assessment for learning in upper elementary grades. *Phi Delta Kappan, 90*(6), 419-421.

Stiggins, R., & Chappuis, J. (2005). Using student-involved classroom assessment to close achievement gaps. *Theory Into Practice, 44*(1), 11-18.

Suhr, D. (2008). *Step your way through Path Analysis.* Paper presented at the Western Users of SAS Software Conference 2008, Universal City, California.

Sung, Y., Chang, K., Chiou, S., & Hou, H. (2005). The design and application of a web-based self- and peer-assessment system. *Computers and Education, 45,* 187-202.

Suurtamm, C., Koch, M., & Arden, A. (2010). Teachers' assessment practices in mathematics: Classrooms in the context of reform. *Assessment in Education: Principles, Policy & Practice, 17*(4), 399-417.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences.* Los Angeles, CA: Sage.

Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of Mixed Methods Research, 1*(1), 77-100.

Thomas, G., Martin, D., & Pleasants, K. (2011). Using self-and peer-assessment to enhance students' future-learning in higher education. *Journal of University Teaching & Learning Practice, 8*(1), 5.

Thum, Y. M. (1997). Hierarchical linear models for multivariate outcomes. *Journal of Educational and Behavioral Statistics, 22*(1), 77-108. doi:10.2307/1165239

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249-276.

U.S. Department of Education (2000). *Before it's too late: A report to the national commission on mathematics and science teaching for the 21st century.* Retrieved from http://www.csu.edu/cerc/researchreports/documents/BeforeItsTooLate2000.pdf

U.S. Department of Education, National Center for Education Statistics. (2013). *The nation's report card: A first look: 2013 mathematics and reading* (NCES Publication No. 2014–451). Retrieved from http://nces.ed.gov/nationsreportcard/subject/publications/main2013/pdf/2014451.pdf

Vidmar, T. (2011). School and the understanding of knowledge between pragmatism and constructivism. *Journal of Contemporary Educational Studies*. Retrieved from http://www.sodobna-pedagogika.net/wp-content/uploads/2013/03/2011_1_eng_02_vidmar_school_and_the_understanding.pdf

Voogt, J., & Kasurien, H. (2005). Finland: Emphasizing development instead of competition and comparison. In Organization of Economic Co-operation and Development, *Formative assessment: Improving learning in secondary classrooms*. Retrieved from http://www.oecd.org/edu/ceri/34260381.pdf

Wagner, M. L., Suh, D. C., & Cruz, S. (2011). Peer- and self-grading compared to faculty grading. *American Journal of Pharmaceutical Education, 75*(7), 130. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3175657/

Wallace, W. V. (2003). *Formative assessment: Benefit for all*. (Master's thesis). Retrieved from http://etd.fcla.edu/CF/CFE0004955/Formative_Assessment_Benefit_For_All.pdf.

Warner, Z. B., Chen, F., & Andrade, H. (2012). Student self-assessment in middle school mathematics: A pilot study. *Northeastern Educational Research Association Conference Proceedings,* Paper 5.

White, B. Y., & Fredrickson, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*(1), 3-118.

Wiliam, D. (1999). Formative assessment in mathematics: Part 2: Feedback. *Equals: Mathematics and Special Education Needs, 5*(3), 8-11.

Wiliam, D. (2007a). *Five "key strategies" for effective formative assessment*. Reston, VA: National Council of Teachers of Mathematics. Retrieved from http://www.nctm.org/news/content.aspx?id=11474

Wiliam, D. (2007b). *What does research say the benefits of formative assessment are?*. Reston, VA: National Council of Teachers of Mathematics. Retrieved from http://www.nctm.org/news/content.aspx?id=11466

Wilson, C., & Deane, F. (2001). Adolescent opinions about reducing help-seeking barriers and increasing appropriate help engagement. *Journal of Educational and Psychological Consultation, 12*(4), 345-364.

Wright, S. (1960). Path coefficients and path regressions: Alternative or complementary concepts? *Biometrics, 16*(2), 189-202.

Zimmerman, B. J. (2001) Theories of self-regulated learning and achievement: An overview and analysis. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed.) (pp. 1-36). Mahwah, NJ: Erlbaum.

Zimmerman, B. J., & Pons, M. M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal, 23*(4), 614-628.

APPENDIX A

Intervention Outline

The instructional sequence will occur over 10 hour-long sessions across approximately 3 weeks.

Session 1

Goal. Students will better appreciate the need for their participation in the assessment process.

Evidence that the goal has been met. Students will submit questions and comments written on exit slips and left on the classroom "Parking Lot" at the conclusion of the session.

Procedure. Session 1 will proceed as follows:

1. For 5 to 10 minutes, students will independently answer a writing prompt, displayed here in Figure A.1.

> 1) Describe a time you completed a task or assignment, at school or outside of school, when you thought you did a good job, but an adult or another student told you that you didn't.
>
> b. Did you agree with the adult or other student's judgment?
>
> - If so, what helped you realize that he/she was correct?
> - If you did not agree with the adult or other student, what makes you think that you were correct?

What information helped you or could have helped you make sure you were doing the best job possible to complete the task?

Figure A.1. Session 1 writing prompt, designed to incite student awareness of past assessment experience.

2. For 5 minutes, students will discuss their answers to the writing prompt with a partner.

3. For 10 minutes, the pairs will use an iPad app, Lino (http://en.linoit.com), that allows them to place virtual post-it notes on a virtual canvas, which will be displayed on the class overhead via computer. Each pair will decide, between their two sets of prompt answers, the best response to each question number and indicate the choice on color-coded post-its. Question 1 will be answered on yellow post-its, question 2 will be answered on blue post-its, and question 3 will be answered on green post-its.

4. For another 5 to 10 minutes, each pair will examine the displayed post-its and add a comment or question about any of their peers' responses on a pink post-it.

5. I will use the students' experiences, comments, and questions to frame the purpose of the upcoming set of instructional sequences: to learn about and practice a set of strategies that will help them better understand what good mathematics work looks like and how they can help themselves and their friends gain a clearer picture of their progress in math, by which they can learn better and understand more in class. Research-based statements that will be highlighted throughout the discussion include the following: (a) Students learn more and perform better on assignments when they practice evaluating work, because they get a better idea of what good work looks like; (b) students frequently help each other identify ways to improve work because they can communicate more easily with each other than with teachers; and (c) the ability to effectively evaluate one's

own work will benefit students not only throughout their school years but also in situations outside of school, such as future jobs.

6. Pairs will complete the session by placing a post-it with a comment or question about the session's activity on the Parking Lot canvas.

Sessions 2 and 3

Sessions 2 and 3 comprise a single activity that requires two class periods to complete.

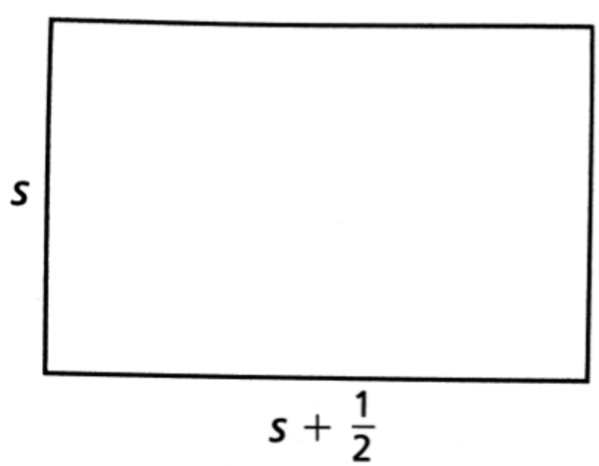
Goal. Students will begin to develop an awareness of their current ability level in identifying characteristics of high-quality work and opportunities for improving their own work, according to assessment criteria.

Evidence that the goal has been met: Students will produce and submit written reflections after they revisit their assessments of two simulated peer responses to an assessment task and compare/contrast their evaluation with one developed by the class as a whole.

Procedure. Sessions 2 and 3 will proceed as follows:

1. Students will be given a blank assessment item, displayed here in Figure A.2.

Use the picture below to answer the question



A. Write an algebraic expression that would represent the perimeter of the rectangle

B. Is your expression above, in Part A, *simplified?* If so, explain how you know. If your expression in Part A is not simplified, simplify it and explain your work.

Figure A.2. Sessions 2 and 3 blank assessment item.

2. Students will be given a constructed response question rubric, displayed here in Figure A.3.

<table>
<tr><td colspan="2">Rubric - Constructed Response Question
Directions: Place a checkmark in the blank next to each indicator that the student's answer has fulfilled

Part A　　　　　　　　　　Part B</td></tr>
<tr><td>____ 1 Point: THE STUDENT completed every part of the question in Section A</td><td>____ 1 Point: THE STUDENT completed every part of the question in Section B</td></tr>
</table>

____1 Point: THE STUDENT correctly answered every part of the question in Section A	____1 Point: THE STUDENT correctly answered every part of the question in Section B
____ 1 Point: THE STUDENT clearly explained every part of his/her answer in Section A with correct words, mathematical symbols, pictures or other evidence.	____ 1 Point: THE STUDENT clearly explained every part of his/her answer in Section B with correct words, mathematical symbols, pictures or other evidence.

Number of points awarded for Part A ___________________

Plus

Number of points awarded for Part B ___________________

Total number of Points for task ___________________

Figure A.3. Sessions 2 and 3 rubric for the constructed response question and point calculator.

3. In groups of 2 or 3 and for approximately 20 minutes, students will discuss possible characteristics of an assessment answer for satisfying each indicator. They will record these ideas on the iPads provided, using the Show Me (http://www.showme.com/learn) or Educreations (https://www.educreations.com) whiteboard app.

4. Groups will present their ideas, and we will assemble a single master list on the whiteboard to be displayed on the overhead.

5. I will distribute two simulated student-work samples of the assessment task. Students will be given two simulated peer responses—designed to represent different levels of quality—to the same assessment item, two copies of the assessment rubric, and a set of highlighters.

6. Students will be asked to use the highlighters to color-code evidence of rubric indicators met in the two student work samples. For example, a rubric indicator might be color-coded orange and the evidence of the met indicator would be highlighted orange within the work sample. They will then assign a score, according to the rubric criteria, for each work sample.

7. Groups will present their ideas and evidence to the class, while I keep a master list. I will facilitate discussion as each group reaches a consensus to generate a whole-group evaluation and scores for the simulated assessment samples; I will address any disagreements.

8. When needed, I will offer students evidence-based, task-oriented feedback addressing their evidence to help them calibrate its accuracy. I will ask students to cite other evidence that could have been included in each student response to make it more effective.

9. The session will conclude when the individual groups revisit their assessments of the simulated tasks and produce and submit a written compare/contrast summary of their evaluation and of the whole-group evaluation.

Session 4

Goal. Students will continue to develop an awareness of their current ability level in identifying characteristics of quality work and opportunities for improving their own work, according to assessment criteria.

Evidence that the goal has been met. Students will produce and submit a written compare/contrast of their own small-group-generated evaluations of assessment task responses with the whole-group-generated one.

Procedure. Session 4 will proceed as follows:

1. Students will be given two simulated student responses to an assessment item, displayed here in Figure A.4.

A compact disc (CD) player costs \$147.35, and CDs cost \$16.95 each (including sales tax).

a. Given that *C* represents the total cost of owning the CD player and the CDs, and *x* represents the number of CDs purchased, write an equation that represents the cost *C* as a function of the number of CDs *x*.

b. Including the cost of the CD player, show how your equation can be used to determine how much someone spent in the first year, purchasing an average of 1.5 CDs per month.

Figure A.4. Session 4 assessment item, to be accompanied by two simulated student responses.

2. Students will be given two copies of the rubric displayed in Figure A.3 above.

3. In pairs, students will evaluate the two responses, citing evidence within the tasks that indicators have been met.

4. Within a whole-group setting, students will discuss their pair evaluations and identify any inconsistencies among groups in order to generate a consensual whole-group evaluation of the samples.

5. I will provide any necessary teacher feedback, which will be task- and evidence-based, on the accuracy and quality of the students' assessments.

6. At the conclusion of the session, student pairs will compile a list of similarities and differences between their own evaluations and that generated through whole-group consensus.

Sessions 5 and 6

Sessions 5 and 6 will involve the same series of steps but with different assessment items for each session. The only difference between the sessions is that, before engaging in the Session 5 task, students will receive written teacher feedback on their Session 4 comparison of their own work with the exemplary sample.

Goal. Students will assess their own ability to generate quality work and opportunities for improving their own work, according to assessment criteria.

Evidence that the goal has been met. Students will produce written reflections in which they compare and contrast their own small-group-generated assessment responses to the whole-group-generated exemplary one.

Procedure. Sessions 5 and 6 will proceed as follows:

1. Students will be given an assessment task, displayed here in Figure A.5.

Coach Wilson ordered T-shirts for the basketball team from two different T-shirt suppliers. One supplier charged $ 16 for each shirt, plus 5% for shipping. The other supplier charged $18 for each shirt, plus 7% for shipping.

A. Coach Wilson ordered the same number of shirts from each supplier. Write two expressions to represent the shipping charges he paid to both suppliers, using x to represent the number of shirts ordered.
B. The first supplier gave coach Wilson a discount of 10% off his order, plus a flat rate of $2 shipping. Write and solve an equation to determine how many t-shirts he bought if the total cost of the order was $146.

Figure A.5. Session 5 assessment task.

Meaghan bought 3 sweaters for the same price and used a coupon for $20 off a total purchase.

A. Write an equation that would represent the situation, if Meaghan spent a total of $100, before tax.
B. How much was each sweater? How much was Meaghan's total if the sales tax was 6%?

Figure A.6. Session 6 assessment item.

2. Student pairs will be given a sample assessment item, on the iPad app Nearpod (http://www.nearpod.com) and copy of the aforementioned rubric. They will be given up to 20 minutes to create, based on the criteria, what they believe to be an exemplary response.

3. While I display the pairs' assessment responses on the class overhead, through the Nearpod app, students as a whole group will examine the work of each pair and identify strengths and opportunities for improvement in each of pair's responses.

4. Students will use specific evidence from the pair-generated assessment samples and whole-group discussion to collaboratively develop a consensual whole-group exemplary response, with any necessary guided, task- and criteria-oriented feedback from me. I will record this exemplary response on the overhead.

5. Student pairs will revisit their own original responses and assemble a list of similarities and differences between the whole-group-generated exemplary response and their own small-group responses.

6. Students will submit this list as an exit slip.

Sessions 7 and 8

Sessions 7 and 8 will involve the same series of steps but with different assessment items for each. The difference between the sessions is that, before engaging in the Session 8 task, students will read my written feedback on their Session 7 comparison of their own work with the exemplary sample.

Procedure. Sessions 7 and 8 will proceed as follows:

1. Before Session 7, students will be given an assessment item, displayed here in Figure A.7. Before Session 8, they will be given a different assessment item, displayed here in Figure A.8. Before Sessions 7 and 8, students will be given an assessment rubric, displayed here in Figure A.9.

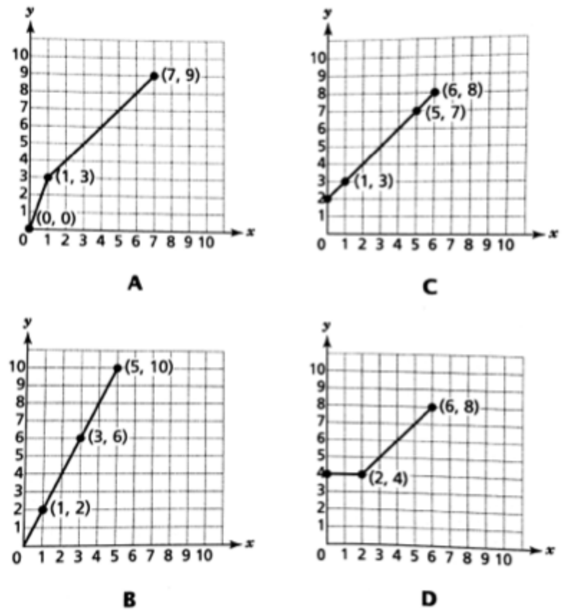
Which graph shows a proportional relationship?  	***Write the correct answer choice here__________** ***Explain, using any words, diagrams, math symbols, or other evidence you would like, why you think the answer is correct.** ***Write one of the incorrect answer choices here__________** ***Explain, using any words, diagrams, math symbols, or other evidence you would like, why you think the answer is incorrect.**

Figure A.7. Session 7 assessment item.

Which table represents a proportional situation?

x	y
25	5
30	6
40	8

A

x	y
25	5
30	10
40	20

C

x	y
25	5
30	4
40	3

B

x	y
25	35
30	40
40	50

D

***Write the correct answer choice here__________**
***Explain, using any words, diagrams, math symbols, or other evidence you would like, why you think the answer is correct.**

***Write one of the incorrect answer choices here__________**
***Explain, using any words, diagrams, math symbols, or other evidence you would like, why you think the answer is incorrect.**

Figure A.8. Session 8 assessment item.

Rubric - Multiple Question

Directions: Place a checkmark in the blank next to each indicator that the student's answer has fulfilled

Correct Answer Selection/Explanation	Incorrect Answer Selection/Explanation
____ 1 Point: THE STUDENT identified the correct multiple choice answer	____ 1 Point: THE STUDENT identified an incorrect multiple choice answer

____ 1 Point: THE STUDENT explained his/her answer with some correct words, mathematical symbols, pictures, or other evidence. OR ____ 2 Points: THE STUDENT completely and clearly explained his/her answer with correct words, mathematical symbols, pictures or other evidence.	____ 1 Point: THE STUDENT explained why the answer was incorrect with some correct words, mathematical symbols, pictures, or other evidence. OR ____ 2 Points: THE STUDENT completely and clearly explained why the answer was incorrect with correct words, mathematical symbols, pictures or other evidence

Number of points awarded for Correct Answer Session __________________

Plus

Number of points awarded for Incorrect Answer Session__________________

Total number of Points for task __________________

Figure A.9. Sessions 7 and 8 rubric for task assessment and score calculator.

2. In pairs, students will complete the assessment task using the rubric as a guide.

3. Students will be given a simulated peer response exemplary of task completion. Using color-coding, students will highlight features common to their own assessment and the exemplary one reflecting that the rubric indicators have been met.

4. Students will note on their own assessment any characteristics of the exemplary sample that are missing from their answer or any characteristics of their answer missing from the exemplary sample.

5. Students will produce and submit a compare/contrast of their own work and the exemplary work.

Sessions 9 and 10

Sessions 9 and 10 will involve the same series of steps but with different assessment items for each session. Before Sessions 9 and 10, students will have the opportunity to review written teacher feedback on the accuracy of their previous revisions.

Goal. Students will use their evaluation skills, developed earlier in the learning sequence, to accurately peer- and self-assess their responses to an assessment task.

Evidence that the goal has been met. Students will produce written reflections about their original peer- and self-assessment revisions.

Procedure. Sessions 9 and 10 will proceed as follows:

1. Before Session 9, students will be given an assessment item, displayed here in Figure A.10. Before Session 10, students will be given a different assessment item, displayed here in Figure A.11. Before Sessions 9 and 10, students will be given peer-/teacher-assessment rubrics and self-assessment rubrics for both the multiple choice question and the constructed response question; these rubrics are displayed here in Figure A.12, Figure A.13, Figure A.14, and Figure A.15.

Which Store has the lowest price?

A) Store A B) Store B C) Store C

D) Store D

CHEDDAR CHEESE

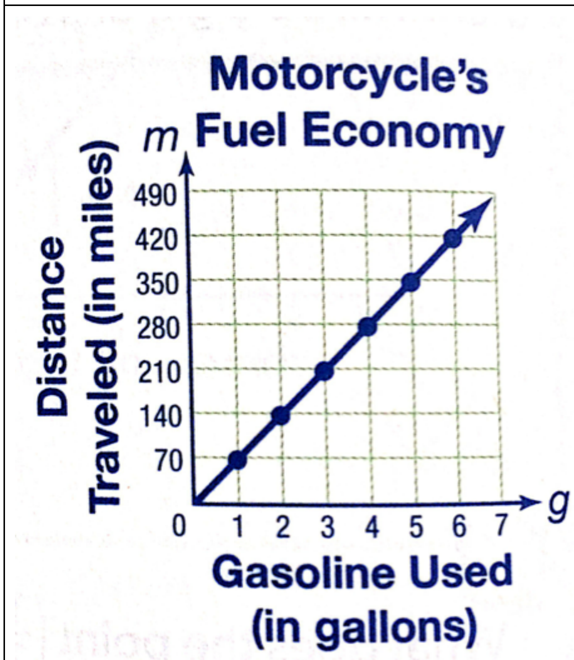
Store	Amount	Price
Store A	3 lb	$9.00
Store B	3 lb	$9.75
Store C	4 lb	$12.40
Store D	5 lb	$14.50

***Write the correct answer choice here__________**

***Explain, using any words, diagrams, math symbols, or other evidence you would like, why you think the answer is correct.**

***Write one of the incorrect answer choices here__________**

***Explain, using any words, diagrams, math symbols, or other evidence you would like, why you think the answer is incorrect.**



A. Does the graph to the left represent a proportional relationship? How do you know?

B. What point on the graph represents the unit rate of the data? How do you know?

Figure A.10. Session 9 assessment item.

The math club needs to raise more than $552.50 for a trip to the state convention. The club has raised 12% of eth funds. Members want to find out how much they each need to raise to make up the difference. Which inequality shows how much money each of the 7 club members needs to raise if each raises the same amount? A. m < $69.45 B. m > $69.45 C. m < $72.35 D. m > $72.35	***Write the correct answer choice here__________** ***Explain, using any words, diagrams, math symbols, or other evidence you would like, why you think the answer is correct.** ***Write one of the incorrect answer choices here__________** ***Explain, using any words, diagrams, math symbols, or other evidence you would like, why you think the answer is incorrect.**

The total charge for a taxi ride includes an initial fee of \$3.25 plus \$2.75 for every ½ mile traveled. Brent took a taxi, which cost him exactly \$17.	**C. Write an equation that represents the situation.** **D. How many miles did Brent travel?**

Figure A.11. Session 10 assessment item.

Teacher/Peer Assessement Rubric - Multiple Choice Question *Directions: Place a checkmark in the blank next to each indicator that the student's answer has fulfilled* Correct Answer Selection/Explanation Incorrect Answer Selection/Explanation	
____ 1 Point: THE STUDENT identified the correct multiple choice answer	____ 1 Point: THE STUDENT identified an incorrect multiple choice answer

____ 1 Point: THE STUDENT explained his/her answer with some correct words, mathematical symbols, pictures, or other evidence. OR ____ 2 Points: THE STUDENT completely and clearly explained his/her answer with correct words, mathematical symbols, pictures or other evidence.	____ 1 Point: THE STUDENT explained why the answer was incorrect with some correct words, mathematical symbols, pictures, or other evidence. OR ____ 2 Points: THE STUDENT completely and clearly explained why the answer was incorrect with correct words, mathematical symbols, pictures or other evidence

Figure A.12. Sessions 9 and 10 teacher-/peer-assessment rubric for the multiple choice question.

Teacher/Peer Assessment Rubric - Constructed Response Question *Directions: Place a checkmark in the blank next to each indicator that the student's answer has fulfilled* Part A	Part B
____ 1 Point: THE STUDENT completed every part of the question in Section A	____ 1 Point: THE STUDENT completed every part of the question in Section B
____1 Point: THE STUDENT correctly answered every part of the question in Section A	____1 Point: THE STUDENT correctly answered every part of the question in Section B
____ 1 Point: THE STUDENT clearly explained every part of his/her answer in Section A with correct words, mathematical symbols, pictures or other evidence.	____ 1 Point: THE STUDENT clearly explained every part of his/her answer in Section B with correct words, mathematical symbols, pictures or other evidence.

Figure A.13. Sessions 9 and 10 teacher-/peer-assessment rubric for constructed response question.

Self-Assessment Rubric - Multiple Choice Question *Directions: Place a checkmark in the blank next to each indicator that the student's answer has fulfilled* Correct Answer Selection/Explanation	Incorrect Answer Selection/Explanation
____ 1 Point: I am certain I identified the correct multiple choice answer OR ____ 1 Point: I think I probably identified the correct answer, but I'm not certain	____ 1 Point: I am certain I identified an incorrect multiple choice answer OR ____ 1 Point: I think I probably identified an incorrect answer, but I'm not certain.
____ 1 Point: I explained my answer with some correct words, mathematical symbols, pictures, or other evidence. OR ____ 2 Points: I completely and clearly explained my answer with correct words, mathematical symbols, pictures or other evidence.	____ 1 Point: I explained why the answer was incorrect with some correct words, mathematical symbols, pictures, or other evidence. OR ____ 2 Points: I completely and clearly explained why the answer was incorrect with correct words, mathematical symbols, pictures or other evidence

Figure A.14. Sessions 9 and 10 self-assessment rubric for the multiple choice question.

Self- Assessment Rubric - Constructed Response Question *Directions: Place a checkmark in the blank next to each indicator that the student's answer has fulfilled* Part A	Part B
____ 1 Point: I completed every part of the question in Section A	____ 1 Point: I completed every part of the question in Section B

____1 Point: I correctly answered every part of the question in Section A	____1 Point: I correctly answered every part of the question in Section B
____ 1 Point: I clearly explained every part of my answer in Section A with correct words, mathematical symbols, pictures or other evidence.	____ 1 Point: I clearly explained every part of my answer in Section B with correct words, mathematical symbols, pictures or other evidence.

Figure A.15. Sessions 9 and 10 self-assessment rubric for constructed response question.

2. Students will independently complete the assessment item, using the rubric as a guide.

3. Students will self-assess their task by citing specific pieces of rubric-based evidence.

4. Students will be given the same simulated peer response to the same assessment item and assess it, also using specific pieces of rubric-based evidence.

5. In a whole-group setting, students will evaluate the simulated peer response, and I will provide any necessary feedback.

6. After this opportunity to receive peer and any necessary teacher feedback, students will revisit their individual assessments and make any adjustments they believe are necessary to improve their accuracy.

7. Students will submit a list of any adjustments they made.

APPENDIX B

Raw Student Data

SID	TEAM	TC	NAPD	PRE-ACH	PRE-SELF	PRE-PEER	POST-ACH	POST-SELF	POST-PEER
1	1	0	1	3	6	2	1	10	4
2	1	0	2	0	10	3	4	2	2
3	1	0	1	1	7	2	2	9	1
4	1	0	1	1	8	1	1	11	6
5	1	0	1	4	0	5	6	0	3
7	1	0	1	3	9	4	2	6	5
8	1	1	2	7	0	4	8	1	3
9	1	1	3	8	3	6	10	2	2
10	1	1	1	1	3	0	5	2	4
11	1	1	1	0	0	6	4	4	1
12	1	1	2	1	0	7	9	1	1
13	1	0	3	6	0	5	5	0	2
15	1	1	2	9	2	3	12	0	0
18	1	1	2	6	0	3	12	0	4
19	1	1	1	2	10	4	5	4	2
20	1	1	2	0	9	1	4	5	0
22	1	1	3	12	7	2	12	2	6
23	1	1	1	2	6	2	8	1	3
24	1	1	1	2	9	1	4	5	1
25	1	1	1	2	1	7	5	3	2
26	1	1	1	1	3	7	1	7	1
28	1	1	3	7	2	2	10	0	0
35	1	0	1	4	4	1	3	4	0
36	1	0	1	1	7	1	5	4	2
39	1	0	1	5	1	4	2	2	2

40	1	1	2	4	6	3	10	2	0
44	1	0	2	1	10	3	1	7	2
45	1	0	1	4	3	2	4	2	5
49	1	0	2	5	7	3	4	6	2
50	0	1	1	4	5	1	7	4	2
51	0	1	1	4	2	0	7	1	2
52	0	1	1	8	1	1	8	0	0
53	0	1	1	5	3	3	6	0	0
54	0	1	1	6	6	6	9	0	6
55	0	1	1	4	0	4	5	2	2
56	0	1	1	4	3	0	7	1	3
58	0	1	2	3	3	4	5	3	2
61	0	0	1	3	8	4	2	7	1
62	0	0	2	0	8	3	2	5	1
63	0	0	1	7	1	2	1	5	0
64	0	0	2	2	6	1	2	4	2
65	0	0	1	1	6	6	0	11	6
66	0	0	2	9	1	3	8	2	1
67	0	0	2	4	4	1	4	6	0
68	0	0	2	1	5	2	6	3	0
69	0	0	2	6	2	1	3	2	2
70	0	0	1	4	3	3	4	6	4
71	0	0	1	2	0	1	2	3	2
72	0	0	1	1	5	2	4	3	1
73	0	0	3	5	3	3	6	0	3
74	0	0	3	9	3	3	7	1	3
75	0	0	3	9	1	1	7	2	3
76	1	0	3	12	3	3	6	6	8
77	1	0	3	5	5	1	7	5	1
78	1	0	2	8	2	3	5	1	1
79	1	0	2	7	4	1	5	5	2

81	1	0	1	3	4	5	3	1	1
84	1	1	2	6	4	6	10	2	1
85	1	1	2	9	1	1	12	1	1
86	0	1	1	2	2	3	6	3	2
87	0	1	2	1	7	4	5	2	1
88	0	1	2	3	5	1	6	8	5
89	0	1	2	5	4	3	7	1	3
90	0	1	3	7	1	2	11	2	1
91	0	1	3	6	4	0	12	0	1
92	0	1	3	7	6	4	12	0	0
94	0	1	3	5	1	3	8	4	0
96	0	1	2	5	5	1	9	1	1
97	0	1	2	7	5	1	7	2	0
98	0	1	1	3	7	2	5	3	0
99	0	1	2	3	6	3	5	2	3
100	0	1	1	3	9	2	6	6	0
104	1	0	2	4	2	3	9	1	3
105	1	0	2	6	1	2	5	1	3
106	0	0	2	5	4	0	7	2	1
107	0	0	1	4	4	1	0	6	0
108	0	0	1	7	1	4	7	3	0
109	0	0	1	2	3	2	3	2	0
110	0	0	2	9	1	3	7	4	2
111	0	0	1	8	1	0	5	6	3

CURRICULUM VITAE

Elizabeth Popelka
1620 Eastern Parkway ● Louisville, Kentucky 40204
(502) 727-2617 ● elizabeth.popelka@jefferson.kyschools.us

Education

PhD Candidate in Curriculum and Instruction, December 2015 (expected graduation)
University of Louisville, Louisville, KY

National Board for Professional Teaching Standards, November 2011

Rank 1 Advanced Practitioner, Middle Grades Mathematics Education, May 2010
University of Louisville, Louisville, KY

Master of Arts in Teaching, Middle School Education, Mathematics, May 2006
University of Louisville, Louisville, KY

Bachelor of Arts, Political Science; Communications (double major), August 1999
University of Louisville, Louisville, KY

Professional Experience

Westport Middle School

July 2014 – Present — Master Lead Teacher of Mathematics; Seventh-Grade Interventionist

August 2013 – June 2014 — Eighth-Grade Mathematics Teacher

August 2012 – August 2013 — Seventh-Grade Mathematics Teacher

South Oldham Middle School

August 2011 – June 2012 — Eighth-Grade Mathematics Teacher; Lead Teacher of Mathematics

August 2008 – August 2011 — Seventh-Grade Mathematics Teacher; Interdisciplinary Team Leader

University of Louisville

August 2011 – December 2011 — Elementary Mathematics Education Part-Time Instructor

Publications

Bush, S., Karp, K., Popelka, E., Miller-Bennett, V., & Nadler, J. (2013). Framing measurement: An art gallery installation. *Mathematics Teaching in the Middle School*, *18*(8), 474-483.

Bush, S., Karp, K, Popelka, E., & Miller-Bennett, V. (2012). What's on your plate? Thinking proportionally. *Mathematics Teaching in the Middle School*, *18*(2), 100-109.

Grant Funding

Universal Design for Learning. (2013). Jefferson County Public Schools ($6,000)

Integrating Technology into Students' Problem-Solving. (2011). Kentucky Council of Teachers of Mathematics ($1,000)

Professional Development Developed and Delivered

Curriculum and Strategies Developed to Promote Student Self-Assessment, Problem-Solving Strategies, and Perseverance through the Use of LiveScribe Smart Pens. (2012, October). Hour-long presentation at the annual conference of Kentucky Council of Teachers of Mathematics. Bowling Green, KY.

Designing and Implementing an Interdisciplinary Mystery Unit. (2011, November). Hour-long roundtable presentation at the National Middle School Association. Louisville, KY.

Designing and Implementing Curricula for Third to Eighth Grade Students at a University of Louisville Summer Camp Integrating Technology, Literacy, Mathematics, and Science. (2009, 2010, 2011, summer). Half-day presentations at the Write One Summer Portfolio Institute. Louisville, KY.

Strategies for Using Various Technological Applications to Formatively Assess and Remediate Students. (2010, August). Half-day presentation to the Oldham County Professional Development Institute. Crestwood, KY.

Use of TurningPoint Technology to Create Engaging Classroom Instruction and Formative Assessments. (2009, August). Half-day presentation to the South Oldham Middle School Professional Development Institute. Crestwood, KY.

Curricula Developed

Seventh-Grade Mathematics Curriculum Frameworks and Key Focus Topics. (2013, spring). Jefferson County public schools. Louisville, KY.

Seventh-Grade Mathematics Common Assessments. (2013, spring). Jefferson County public schools, Louisville, KY.

Middle School Mathematics Curriculum Frameworks. (2011-2012). Oldham County Public Schools. Crestwood, KY.

Middle School Mathematics Transfer Tasks and Rubrics. (2011-2012). Oldham County Public Schools. Crestwood, KY.

Middle School Mathematics Constructed Response Items and Rubrics. (2009). Oldham County Public Schools. Crestwood, KY.

Professional Memberships

Greater Louisville Council of Teachers of Mathematics
Kentucky Council of Teachers of Mathematics
National Council of Teachers of Mathematics
Association for Middle Level Education

Awards and Honors

2011 South Oldham Middle School Staff & Teachers Achieving Results (STAR) Award

2006 University of Louisville Faculty-Selected Outstanding MAT Student in Middle Grades Mathematics

1995 University of Louisville McConnell Scholar for Political Leadership

1995 University of Louisville Presidential Scholar