Highly Effective Instructional Practices in High School Mathematics Classes

by

Piper I. Stratton

Doctoral Study Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Education
School Psychology

National-Louis University
September 2014

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Abstract

Highly Effective Instructional Practices in High School Mathematics Classes

by

Piper I. Stratton

Doctoral Study Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Education
School Psychology

National-Louis University
September 2014

Abstract

Over 40 years of research, three categories of instructional practices are consistently shown to enhance student achievement, including (a) Evidence-based (EB) Strategy Instruction, (b) Feedback, and (c) Formative Assessment. It was the hypothesis of this study that Grade 9 Algebra 1 classrooms do not routinely use these EB practices to enhance their instruction. Data was collected from 12 Algebra 1 classrooms utilizing a researcher developed systematic observation tool featuring highly effective instructional practices from the 2001 Marzano, Pickering, and Pollack and 2009 Hattie meta-analyses. Study results suggested that the frequency of EB instructional practices varied remarkably among teachers. However, the preponderance of teaching time was spent in two forms of practice with little time devoted to other EB strategies and informal formative assessment practices often lacked variety and depth. Last, the frequency or type of EB instructional practices used did not differ between classes designed for students with average math skills compared to classes designed for lower skilled students. Recommended methods for increasing the widespread use of highly effective EB instructional practices included: (a) system-wide improvements in pre-service teacher training in highly effective instructional practices, (b) more effective on-the-job professional development and implementation practices, and (c) the use of structured professional learning communities focused on improving pedagogy.

Acknowledgments

First and foremost, I would like to thank my family for putting up with me while I returned to school and took on this overwhelming project. I could not have done it without the support of my husband, Mike Gara, who encouraged me and reminded me that it would be over someday; and my kids who mourned the loss of Saturday morning crepes, but were gracious enough to occasionally say, “You did warn us.”

I would also like to thank my dissertation chairperson, Dr. Mark Shinn, who stayed with me throughout this project despite his many commitments. I could not have sustained the momentum to ever finish this without him.

A big thank you to Dissertation Committee members, Dr. Jan Perney, Dr. Daniel Newman, and Dr. Diane Salmon for agreeing to work with me. Thanks for your time, knowledge, and feedback.

Last, but not least, I want to acknowledge my mother, Dawn Atwater-Haight, for her unwavering faith in all three of her children. You instilled in us your belief in limitless possibility, devotion to fulfilling our goals even in the face of struggle, and an acceptance of our imperfections. You are truly nothing short of inspirational.

Table of Contents

List of Tables

List of Figures

Highly Effective Instructional Practices in High School Mathematics Classes

CHAPTER 1- STATEMENT OF THE PROBLEM

A 2010 press release by U.S. Department of Education Secretary Arne Duncan stated:

> More parents, teachers, and leaders need to recognize the reality that other high-achieving nations are both out-educating us and out-competing us. Being average in reading and science- and below average in math- is not nearly good enough in a knowledge economy where scientific and technological literacy is so central to sustaining innovation and international competitiveness (U.S. Department of Education, 2010).

Secretary Duncan's concerns about the implications of below average academic performance and its economic impacts were echoed by Byron Auguste, one of the contributors to the 2009 McKinsey and Company report, *The Economic Impact of the Achievement Gap in America's Schools.* He lamented that American students, on the whole, have no idea how far behind they are academically with respect to a job market that is highly competitive internationally. He agreed with the common American assumption that lagging achievement is a problem for our economically disadvantaged and minority students, but went on to say, "The achievement gap not only effects poor children in failing schools, but most children in most schools"(McKinsey on Society [videocast], 2010).

Achievement is so consistently low that it also shuts many U.S. students out of the military. According to a recent study of almost 350,000 young adults with a high school diploma who took the Armed Services Vocational Aptitude Battery (ASVAB), almost a quarter (23%) failed to earn the minimum score to qualify for enlistment (Theokas, 2010). The ASVAB assesses abilities considered foundational for a comprehensive range of occupations available in the military, and is the most frequently used test of its kind in the world. Those persons who fall short on the ASVAB are not only considered unprepared for military jobs, they are likely unprepared for many civilian jobs. Unsatisfactory ASVAB scores were highest for minority young adults, with 29% of Hispanics and 39%of African Americans judged ineligible for military services based on their low scores.

> This shatters the comfortable myth that academically underprepared students will find in the military a second-chance pathway to success. For too long, we educators have dismissed worries about the low academic achievement of 'those students' with the thought that 'if they're not prepared for college or career, a stint in the service will do 'em some good' (Theokas, 2010, p. 1).

Joining the military is therefore not a viable back-up plan for many young adults with low skills.

Possible Causes of Poor Skills and the American Achievement Gap

Given the complexity of teaching and learning in schools, researchers are understandably reluctant to do more than conjecture about the probable causes of poor skills and the international achievement gap. However, among many plausible explanations for this state of affairs, at least two are prominent. First, too many students have not mastered the basic skills in reading and math from one year to the next, resulting in a lack of readiness to move on to higher-level concepts, culminating in a large basic skills gap by the time they reach high school. Second, once in the high school door, their teachers may not be using the most powerful evidence-based curriculum and pedagogical (i.e., instructional) practices.

First, the reality that US high school students' overall academic performance does not measure up internationally and that the situation in math is even worse is established. Next, the evidence of poor US math skills is confirmed based on national testing data indicating that even the average Grade 8 student lacks readiness for high school algebra, the gateway to all higher-level mathematics course-taking and post-secondary success (Schiller & Muller, 2003). Generally poor math skills paired with the established importance of Algebra 1 is linked to a key questions about typical US Algebra 1 curriculum and instruction effectiveness: Do teachers know and use the most powerful instructional methods available?

U.S. Students' Academic Performance Does Not Measure Up Internationally

The preponderance of evidence indicates that the United States' education system is failing to prepare our young people for jobs or careers to sustain themselves financially, much less to compete in the global economy with better-educated persons from other countries. It has long been the case that students in

America's high schools are struggling compared to international achievement standards. The Program for International Student Assessment's (PISA) finds that US 15-year-olds continue to perform only around the international average in reading and science with the most current rankings as 14th and 17th respectively out of 34 OECD countries (Organisation for Economic Cooperation and Development, 2011). The PISA assessments are administered every 3 years and are designed to determine what students know and whether students can apply their knowledge and acquired skills to the real world. Almost one in five of US 15-year-olds did not reach the 2009 PISA baseline *Level 2* in both reading and another 20% fell below the same level in science. *Level 2* represents a minimum capacity for basic tasks like locating information, making personal connections with text, or in science, "demonstrating the science competencies that will enable them to participate effectively and productively in life situations related to science and technology" (Organisation for Economic Cooperation and Development, 2007, p. 44).

U.S. mathematics international performance is especially problematic. If average international performance on the PISA reading and science tests does not bode well for the country's international economic competitiveness, the situation in mathematics is far worse. On the 2009 PISA, United States 15 year-olds ranked 25th out of 34 countries, placing them in the bottom one-third of OECD countries. Among the countries with average PISA math scores higher than the U.S. were countries like Korea, Finland, Switzerland, and Japan, and the

U.S. was surprisingly also out-performed by smaller countries like the Slovak Republic, Slovenia, Iceland, and Estonia. Close neighbor of the U.S., Canada, ranked 5th among all OECD countries (Organisation for Economic Cooperation and Development, 2011).

Problematically, compared to high performing countries, the U.S. education system turns out more students with the lowest level of PISA math skills. Almost one in four (23.4%) U.S. high school students scored below PISA *Level 2*, a level of mathematics proficiency defined as "a baseline level of mathematics proficiency at which students begin to demonstrate the kind of literacy skills that enable them to actively use mathematics"(Organisation for Economic Cooperation and Development, 2004, p. 54). This percentage is significantly higher than the OECD country average of 20.8% of students falling below *Level 2*. In other words, the U.S. turns out more students with very poor skills who will struggle to compete in the global job market.

Conversely, the U.S. also turns out fewer students with the highest level of math ability compared to high performing countries (OECD, 2011). Only 12% of U.S. students scored at the highest levels, *Levels* 5 and 6, compared to the international average of 16%. As a striking point of comparison, 50% of students in Shanghai-China reach at least Level 5, while 30% and 20% of students reached the same benchmark in Singapore and Hong Kong-China, respectively. Chinese Taipei, Korea, Switzerland, Finland, Japan and Belgium also produced 20% of students at the top levels of achievement. The U.S. is therefore producing far fewer students prepared to compete at the top levels of the highly technological global job market; jobs that have an important impact on the U.S. economy.

Poor U.S. mathematics performance not unique to PISA scores. The Trends in International Mathematics and Science Study (TIMSS) is an international comparative study of educational achievement in mathematics and science developed by the International Association for the Evaluation of Educational Achievement (IEA). The purpose of TIMMS is to identify international trends in mathematics and science achievement at Grades 4, 8, and 12 for the purpose of providing guidance to educational policymakers, administrators, teachers, and researchers (Mullis, Martin, Foy, & Arora, 2012). TIMSS 2011 represents the fifth cycle of the assessment, which has been administered every 4 years since 1995. A focus on assessing the acquisition of skills aligned with curriculum represents the major difference of TIMMS compared to PISA, the latter of which targets skills students acquire and can apply to real-world settings (National Center for Education Statistics, 2004). In contrast to the PISA, which is a one-time, single grade/age test (i.e., 15-year-olds), TIMMS tests the same students at Grade 4, 8, and 12 also allowing for an examination of inter-individual, longitudinal trends.

TIMMS 2011 mathematics results by benchmark and country are shown in Figure 1. According to TIMMS 2011, U.S. eighth graders had an average score of 509 in mathematics, within the international average range (500). Although this average score suggests better international mathematics performance than the

PISA, still, almost one in three (32%) U.S. students performed at or below the *Low* benchmark. Performance at or below the *Low* benchmark indicates these students had a grasp of only the most basic mathematical concepts. By contrast, Chinese countries had less than one in eight students (12%) who performed at or below the *Low* benchmark. At the highest level of achievement, *Advanced*, TIMMS 2011 data were consistent with PISA results. Few U.S. students did well; *TIMMS* showed 7% and *PISA* showed 10% of U.S. eighth graders met the high benchmarks. Again the top-performing PISA countries produced impressive TIMMS results. For example, Chinese-Taipei, Singapore, and Korea had between 47% and 49% of their eighth graders meet the *Advanced* benchmark, seven times more students than in the U.S.

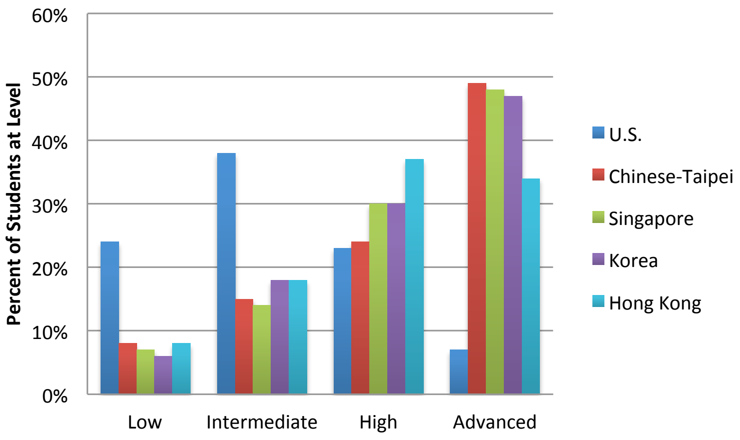

Figure 1. Percentage of eighth graders meeting TIMMS math benchmarks in the U.S. and high-performing comparison countries.

National and State Achievement Data Consistent with International Data

If international data are not concerning enough, conclusions about U.S. students' mathematics skills are echoed by results from the U.S. Department of Education's National Center for Education Statistics (NCES). The National Assessment of Educational Progress (NAEP, 2011) is the largest national and ongoing assessment of American students' knowledge and skills in the subject areas of reading and math at Grades 4, 8, and 12. NAEP math assessment results are reported at three achievement levels, *Basic*, *Proficient*, and *Advanced* (National Center for Education Statistics, 2011). At Grade 8, math performance at the *Basic* level "signifies an understanding of arithmetic operations including estimation of whole numbers, decimals, fractions, and percents" (National Center for Education Statistics, 2011). To be considered *Proficient*, a Grade 8 student should be able to defend their ideas, and give supporting examples, understand the connections between fractions, percents, decimals, and other mathematical topics such as algebra and functions, and the ability to problem solve in practical situations (National Center for Education Statistics, 2011). Not surprisingly, 2011 NAEP results are consistent with PISA and TIMSS assessment outcomes, confirming that (a) many American students fail to attain a foundational level of proficiency in math, and (b) not enough students reach the top levels of proficiency.

Local state data consistent with national averages. The NAEP data also indicate that Illinois' student achievement is as poor as the national condition.

Illinois ranked 22 out of 36 states on the NAEP Grade 8 mathematics results. Two-thirds (67%) of Illinois eighth graders demonstrated math skills below Proficiency. The percentages of U.S. students meeting each benchmark in mathematics for the 2009 PISA, 2011 TIMMS, and 2011 NAEP National and Illinois tests are shown in Table 1. The agreement in scores between NAEP, PISA, and TIMMS is especially clear at the lowest and the highest levels of achievement with few students reaching advanced proficiency standards and nearly a quarter of U.S. students across all three tests struggling to meet the lowest benchmark.

Table 1

Percentages of U.S. eighth-grade students reaching benchmarks across the 2009 PISA, 2011 TIMMS, and 2011 NAEP.

Test/ Benchmark Equivalent	PISA 2009	TIMMS 2011	Overall NAEP 2011	Illinois NAEP 2011
Advanced	10	7	8	8
Intermediate/ Proficient	42	24	27	25
Basic	24	38	38	40
Below Basic	22	24	27	27

Even fewer poor and minority students reach proficiency. As much as the educational fate of the typical American student is concerning, it should not be overlooked that the problem is worse for several sub-groups, such as students from poor, ethnic, and racial minority backgrounds. According to the National Center for Education Statistics (2011) a 10-point difference on the NAEP is roughly the equivalent of 1 year of school. On the 2011 Illinois NAEP, African-American students had an average score that was 33 points lower than their White Illinois counterparts, equivalent to being almost 3 years behind. African-American students made up one-third of the students in the lowest-performing quartile. Hispanic students had an average score that was 22 points lower than White students, equivalent to almost 2 years of learning, and comprised another one-third of the bottom quartile. In addition, English-Language Learners (ELLs), those limited English-speaking students new to the U.S., made up a full 15% of the lowest performers. Regardless of ethnic status, low-income students (i.e., those eligible for free/reduced-price school lunch) also achieved at significantly lower levels than average students. Low-income students had an average score that was 27 points lower than students who were not eligible for free/reduced-price school lunch. In fact, more than two out of three students performing in the bottom quartile came from low-income families. By Grade 8, these groups, on average, have fallen between 2 to 3 years behind typical students.

Is U.S. Math Achievement Improving, and Is It Enough?

Over time, TIMMS data indicate that American eighth graders have made small yet statistically significant gains over the course of the 8 years between 2003 and 2011. However, the improvement is still not closing the gap between the U.S. and top-performing countries. A close look at the TIMSS data reveals the insufficiency of U.S. gains through 2011. For example, the average score gain for American eighth-graders between 2003 and 2011 was less than a one-point increase in achievement. This gain is far lower than the 16 point average gain for South Korean students in that same period (Biddle, 2012), though U.S. students started out lower and had more room to grow. In short, despite gains, the gap in achievement between the U.S. and top performing countries has gotten larger, not smaller.

On other measures, U.S. scores have not changed significantly. PISA assessment scores remain static in the periods from 2000, 2003, and 2006 (Organisation for Economic Cooperation and Development, 2011). NAEP reported only slightly more significant math gains for fourth and eighth graders according to trends since the test's inception in 1973 (National Center for Education Statistics, 2011). The question remains whether these small gains are meaningful compared to that of other countries. To be able to compare the progress of U.S. students to student progress in other countries, Harvard's Program on Education Policy and Governance and Education Next created a common metric across the PISA, TIMMS, and NAEP. Although multiple academic domains were evaluated, the picture for math alone is similar to the combined results. Averaged across all three tests, U.S. students' test-score performance increased by 1.6% of a standard deviation per year over the period between 1995 and 2009, or approximately 22% of a standard deviation total, the equivalent of 1 year's worth of learning (Hanushek, Peterson, & Woessmann, 2012). Although U.S. gains were evident, in the same period other countries made similar or even more impressive gains not explained by "catch-up theory," the theory that countries that start out with lower scores have more room for growth. "While 24 countries trail the U.S. rate of improvement, another 24 countries appear to be improving at a faster rate. Nor is U.S. progress sufficiently rapid to allow it to catch up with the leaders of the industrialized world" (Hanushek et al., 2012, p. vi). Of further concern is the fact that although gains were observed in lower grades (e.g. Grades 4 and 8) they did not translate into improved high school performance. "Students themselves and the United States as a whole benefit from improved performance in the early grades only if that translates into measurably higher skills at the end of school"(Hanushek et al., 2012, p. vi). The ultimate failure of an educational system is perhaps most evident at the point at which the student exits formal schooling at the end of high school.

The Importance of Algebra for High School Mathematics Success and Beyond

The national and international testing results clearly demonstrate that many U.S. students fail to master foundational mathematics skills emphasized in elementary and middle school curriculum. This lack of prerequisite knowledge and skills serves as a significant barrier to further mathematics learning, as students' new learning depends on that prior knowledge: "Every new thing that a person learns must be attached to what the person already knows" (McLaughlin et al., 2005, p. 5). Even average eighth graders have relatively weak prerequisite skills such as interpreting symbols, completing tables, finding sums of series, generalizing patterns, or solving word problems (Loveless, 2013), and the bottom 25% of learners are of even greater concern. Students with skills in this range have significant skill gaps, such as a lack of facility with fractions, making success in algebra unlikely. Wu (2001) and the National Mathematics Advisory Panel (2006) specifically attribute the failure of many students to master higher-level math, starting with algebra, to the pervasive prior lack of deep understanding of whole numbers and fractions. Too many students have not mastered the basic skills from one year to the next, resulting in a lack of readiness to move on to higher-level concepts, culminating in a large basic skills gap for many students by the time they reach high school. Students therefore enter high school expected to learn algebra without the necessary foundations of fluency with whole numbers and fractions, measurement, and geometry.

Why is algebra important? In the last two decades the importance of requiring higher-level course work, such as Algebra 1, Geometry, and Algebra 2 for all students has been emphasized for three main reasons. First, participation in more and higher-level mathematics courses is correlated with higher mathematics achievement, regardless of background factors (Gamoran, 1987; Hoffer, Rasinski, & Moore, 1995; Ma, 2001). Second, Algebra 1, distinct from elementary math with the introduction of unknown variables, has been named the gatekeeper for all higher-level mathematics course taking (Schmidt, Rotberg, & Siegel, 2003). Third, successful completion of these courses has been shown to correlated with post high school success (Schiller & Muller, 2003). For these reasons, some school reform activists have declared access to algebra a "new civil right" (Jetter, 1993; Moses, 1995), considering the requirement of algebra-for-all a possible equalizer for traditionally under-performing groups. Research on post-secondary success and the resulting sentiments regarding racial equality catalyzed increases in math course requirements across America's high schools from 2 years mandatory math to 3 years including Algebra 1, Geometry, and Algebra 2 as the minimum (Schmidt et al, 2003).

The algebra dilemma. The pervasive U.S. student lack of algebra readiness by high school paired with the established importance of Algebra 1 creates a dilemma for American secondary educators. It is incumbent on Grade 9

Algebra 1 teachers to provide curriculum and instruction that addresses skill gaps and prepares students for even more challenging course work in the future. The curriculum and instruction at the high school level must therefore be highly effective.

Failure to Employ "What Works"

Despite the recognized need to make up for achievement gaps and prepare students for their futures, Algebra 1 teachers may not be using the most powerful evidence-based curriculum and pedagogical (i.e., instructional) practices. Early in the 21st century, the U.S. government began to respond to the low mathematics performance of students on international and national math tests and demanded a large-scale review of the available quality research on math curriculum, instruction, and practices. The analysis led to conclusions that both the U.S. math curriculum and instructional practices were to blame for the underachievement of American students after Grade 4 (Kilpatrick, Swafford, & Findell, 2001).

Curriculum contributions. For the last three decades, the major influence on math curriculum was a 1989 report by the National Council of Teachers of Mathematics (NCTM), *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989). To the dismay of many educational researchers, the NCTM standards were not based on experimental or quasi-experimental research, but were theory driven. Field-testing was suggested only *after* curricular changes were in place (Carnine & Gersten, 2000). Despite the lack of evidential basis for the recommendations, the NCTM document undergirded much of the curriculum built after 1989 and continues today. Critics suggest that the NCTM standards resulted in curriculum considered “shallow, undemanding, and diffuse in content coverage” (Kilpatrick et al., 2001, p. 36) and excessively large textbooks, covering too many topics, repetitively, with too little depth (Valverde & Schmidt, 2007).

Inconsistent instruction. As stated by Kilpatrick et al. (2001) “even with high standards, exemplary textbooks, and powerful assessments, what really matters for mathematics learning are the interactions that take place in the classrooms" (p. 45). Observational studies suggest that these interactions have changed little in the last half-century. Math instruction in U.S. schools is characterized as recitation teaching, which is composed of teacher lecture interspersed with questions and brief student responses with simple teacher acknowledgment indicating if the student is right or wrong. This pattern of question, response, and acknowledgment repeats until the material for the day is covered and then independent work is assigned (Kilpatrick, 2001). Two important questions must be asked to evaluate the efficacy of typical instructional practices:

1. What are high-quality, evidence-based instructional practices?, and
2. Is typical U.S. math instruction comprised of these high-quality, evidence-based practices?

For the purpose of this study, without denying the critical importance of a coherent, well-designed curriculum with high standards for mastery, the focus is limited to a discussion of the lack of, or inconsistent use of high-quality, research-based instructional practices in U.S. schools?

High Quality Instructional Practices

The quest for the "holy grail" of instructional practice, teacher pedagogical practices that are superior in increasing student achievement, is not new. For the past 40 years, educational researchers have examined the vast database of research on instructional practices to pinpoint those actions that deliver the best achievement outcomes for all students (e.g., Berliner (1984); Gage (1984) Walberg (1984), Wittrock (1984), Brophy and Good (1986) Rosenshine and Stevens (1986), Marzano, Pickering, and Pollack (2001), Marzano (2007), Hattie (2009), Cornett (2010), among others). These effective teaching practices syntheses have accomplished at least two important objectives. First, they de-bunked the prevailing sentiment prior to the 1970s that factors outside of school (e.g. socio-economic status, parenting) were so influential that teachers could do little to alter learner outcomes. What the teacher does *is* related to student achievement. Second, the syntheses showed that despite the acknowledged complexity of teaching, several instructional strategies exist that reliably enhance student achievement. Synthesis after synthesis revealed the effectiveness of a reasonable number of teaching practices that, if used consistently, have the potential to raise the achievement of U.S. students significantly.

A typical example of some of the early consensus about such practices can be found in the Rosenshine and Stevens (1986) synthesis completed almost 30 years ago. These researchers identified six "fundamental instructional 'functions'" (p. 179) that enhanced student achievement across grade levels and subject areas and for both basic and advanced skills.

1. Review prerequisite knowledge (e.g., check and reteach previous day's work),
2. Present new content/skill with clear modeling of procedures,
3. Guide student practice with checks for understanding,
4. Provide feedback and correctives, with re-teaching, if needed,
5. Allocate time for independent student practice with monitoring, and
6. Provide weekly and monthly reviews.

Marzano and others (Marzano, 2007; Marzano et al., 2001) identified teaching practices similar to those of Rosenshine and Stevens (1986) in their own subsequent syntheses, adding practices like the strategic use of Visual and Graphic Devices (i.e., Non-linguistic Representations) and Problem-Solving Strategy instruction (i.e., Generating and Testing Hypotheses) to the list of effective instructional practices.

The latest synthesis of instructional practices that enhance achievement comes from Hattie (2009). In his comprehensive study of over 800 meta-analyses relating to student achievement, he articulated the importance of recognizing the plethora of “evidence-based” solutions that have come out of the education literature, resulting in the sense that “everything works”(p.6). Hattie found that almost every study he investigated indicated some degree of positive influence on achievement, so he determined that .40, the average effect size for all of the educational influences, should be considered the hinge-point, rather than zero. Therefore, only instructional practices with effect sizes greater than .40 are understood to have a better than average influence on achievement and that practices meeting this criteria deserve the focused attention of educators. Furthermore, Hattie suggested that instructional practices with effect sizes greater than .60 are excellent. Based on this metric, among the most powerful instructional practices identified by Hattie were Feedback, Frequent Formative Assessment, Meta-cognitive Strategies, Direct Instruction, Mastery Goals and Visual Organizers.

Effective mathematics instructional practices for all. The identification of generic effective instructional practices has proceeded with a simultaneous search for content-specific instructional practices, including mathematics. As part of the mission to identify how to improve our country’s poor mathematics outcomes, a targeted effort to investigate the causes and solutions to America’s poor math performance was made by President Bush who commissioned the National Mathematics Advisory Panel (NMAP, 2006). The NMAP’s Task Force on Instructional Practices asked the question “What instructional practices enable students to learn mathematics most successfully?” (p. 6-xiii). Like Hattie, the

NMAP determined that it is especially important for teachers to actively monitor student understanding and mathematical abilities, and recommended an approach that systematically utilizes Hattie-identified practices such as Frequent Formative Assessment, Feedback, and Mastery Learning. The NMAP also recommended a combination of Direct Instruction and Problem-Based Learning.

Studies specific to mathematics teaching practices echo the findings of summaries of generally effective instructional practices. A meta-analysis by Marcucci (1980) marked a starting place for building a high quality math instruction knowledge base. From the 33 identified studies (11 focusing on Algebra) conducted between 1950 and 1980, Marcucci classified instructional practices into four categories, including Modeling (use of visual aids or manipulative materials) Systematic (a prescriptive approach to problem solving), Heuristic (the teaching of problem solving skills using diagrams or simplifying using smaller numbers), and Guided Discovery (questioning strategies employed to guide students discover solutions). Marcucci found that the Heuristic instructional method was the most effective for improving achievement in mathematics, but positive effects were significant only at the elementary level. At the secondary level (7-12), the Systematic approach, comparable to Direct Instruction methods, was the only grouping with significant positive effects.

Building on Marcucci (1980), Haas (2005) conducted a meta-analysis of 35 studies conducted from 1980 to 2002 at the secondary level where Algebra was the focus. Haas identified six groups of instructional methods: Cooperative

Learning (students working together towards common goal); Communication and Study Skills (teaching students to read and study math effectively and express math ideas verbally or in writing); Technology-Aided Instruction; Problem-Based Learning (teaching deductive and inductive reasoning for problem solving); Manipulatives, Models, and Multiple Representations (teaching concepts using concrete or symbolic representations); and Direct Instruction (explicit review of previously learned content and the sequential teaching of steps, paired with practice and feedback). Direct Instruction (DI) had the highest effect size (ES=.55) regardless of student ability. Problem-Based Learning (PBL) also showed a high effect size (ES=.52), but had medium effects with the lower ability classes studied with a small negative effect on higher ability students. Manipulatives, Models and Multiple Representations had a medium effect size (.38).

More recently, Rakes et al. (2010) conducted a meta-analysis to determine what are the most effective instructional methods used in teaching Algebra at the secondary level, and furthermore, what are the characteristics of Algebra instruction that make the biggest difference in achievement. Rakes et al. included studies from 1968 to 2008, searched 20 electronic databases and out of 594 potentially relevant studies retained 82 studies that met strict inclusion criteria. Instructional Strategies (ES=.35) and Manipulatives (ES=.34) outperformed most curricular factors in enhancing achievement. Instructional Strategies was made up of four individual strategies including Cooperative Learning, Mastery

Learning, Multiple Representations (i.e. Visual and Graphic Devices), and Formative Assessment strategies. Although each meta-analysis grouped variables slightly differently, notably, the instructional practice that proved to be most effective were, in essence, the same as Hattie's (2009) generic (i.e., not subject-specific) strategies, with the clear emphasis on Strategy Instruction and the systematic and frequent use of Formative Assessment to inform instruction and provide Feedback to individuals and groups of students.

Furthermore, syntheses of effective math instructional practices for low-performing students reveal that the same techniques that work for students without an identified math learning disability (LD) also improve learning outcomes for students with identified math LDs (Baker, Gersten, & Lee, 2002; R. Gersten, Chard, Jayanthi, & Baker, 2006; Kroesbergen & Van Luitt, 2003). For example, instructional practices recommended for students with a math LD are so similar to methods already mentioned that they are redundant, including Systematic and Explicit Instruction, Visual and Graphic Organizers, Metacognitive Strategy Instruction, and frequent Formative Assessment with Feedback (R. Gersten & Clarke, 2013; Lloyd, Forness, & Kavale.K.A., 1998; NMAP, 2006). Meeting the needs of lower-performing students may, therefore, require less innovation and more focus on effective instruction known to work for *all* students.

Explicitness: Key Factor of Effective Instruction

One fundamental theme across studies of effective instruction is that more explicit teaching practices increase student learning. In fact, all of the top-ranked instructional practices shown to consistently enhance student achievement across subject areas and student ability levels share explicitness as a common feature. Teacher clarity, a synonym for explicitness, is defined by Hattie as organization, explanation, examples, and guided practice, and was ranked in the top five most effective instructional practices with an overall effect size of 0.75. Archer and Hughes (2011) define explicit instruction as instruction that is "unambiguous and direct" (p.1) involving clear statements about purpose, explanations and demonstrations of the learning target, and practice with feedback until mastery is achieved.

For example, frequent Formative Assessment, Hattie's (2009) number one ranked instructional practice with an effect size of 0.90, requires ongoing and systematic checks of student understanding to enable modification of instruction, as needed. Additionally, corrective Feedback, with an effect size of 0.73, serves the purpose of providing students explicit information to clarify and attain learning objectives. Practice, especially deliberative (i.e., spaced) practice, also ranked very highly among evidence-based instructional practices, with an overall effect-size of 0.71. Deliberative or targeted practice assumes appropriate levels of feedback and is, by definition, more explicit in terms of the path towards the learning objective. Last, Strategy Instruction in general, with an effect size of 0.60 and Meta-cognitive Strategies, with an effect size of 0.69, also ranked very highly in both Hattie's and Marzano's (2001) meta-analyses. Highly effective Strategy Instruction has a stated learning objective and is achieved through explicit modeling, timely cueing (feedback), and deliberative practice.

Teaching practices may be considered to vary along an explicitness continuum from implicit to explicit. Implicit instruction is typified by the presentation of examples or illustrations of content from which students are required to uncover the underlying principles or rules with little teacher guidance. Problem-Based Learning (PBL) that emphasizes inductive reasoning is an example of implicit instruction and is characterized by student-driven analysis of real-world problems. The problems themselves are used as tools to attain the required knowledge and the skills necessary to eventually solve the problem (Gijbels, Dochy, Van den Bossche, & Segers, 2005). PBL and similar implicit instructional methods, also called Inquiry Learning or Inductive Reasoning, albeit not ineffective, are like most instructional strategies, with only low to average effect sizes from 0.15 to 0.33 (Hattie, 2009; Marzano et al., 2001). Consequently, although implicit methods can have positive effects on student achievement, they do not rank among the most powerful instructional methods available to teachers.

As instructional techniques gain in explicitness, they also gain in effectiveness. For example, Problem Solving Teaching, an instructional method where students are explicitly taught a heuristic or framework often supplemented with visual representations with which to solve problems, has a more powerful effect size of .61 (Hattie, 2009). The more explicit Problem Solving Teaching method, therefore, doubles the achievement gains produced by PBL.

The achievement outcomes of Simple Feedback compared to Elaborated Feedback further support the claim that the degree of explicitness of practice is related to increases in student achievement. Specifically, Simple Feedback, which tells the student only what is right or wrong has an effect size of .22 (Bangert-Downs, Kulik, Kulik, & Morgan, 1991), and is therefore much less effective than Elaborated (often corrective) Feedback, with effect sizes at and above .73 (Hattie, 2009) and .90 (Marzano et al.,2001). Corrective or Elaborated Feedback provides the student with a more detailed and explicit explanation of the problem solving process.

Specific Explicit and Effective Instructional Practices That Make a Difference

It can be concluded, therefore, that the most effective mathematics instruction is the most explicit mathematics instruction. Explicit and effective instruction is characterized by practices that fall into a few non-orthogonal categories including (a) Evidence-Based (EB) Strategy Instruction, (b) Feedback, and (c) Formative assessment.

EB Strategy Instruction. Direct Instruction (DI) is an approach to explicit instruction originally developed by Engelmann in the 1960s for very young children (Engelmann & Bereiter, 1966). Over the next 40 years DI programs have been implemented with children of all ages and across subject areas and DI consistently outperforms other instructional methods (Barbash, 2011). DI methods have been shown to be highly effective with elementary and high school students in general education (.99) and special education (.86) (Adams & Englemann, 1996). Most recently, Hattie found DI methods across four meta-analyses had an effect size of .59. DI programs are characterized by the focus on mastery learning via a sequence of seven explicit instructional practices that include goal clarification, explanation of success criteria, explicit modeling, checking for understanding, guided practice, independent practice, and review. In DI, explicitness is often ensured by the use of scripts, teacher modeling, and group responding, especially with respect to initially modeling of the procedures involved in learning the new strategy. Sufficient guided practice followed by independent practice of the new skill is provided and strategies and skills are reviewed to ensure they are learned and generalized within days and weeks of the new learning.

Effective EB Strategy Instruction is characterized by a clear and explicit shift from the teaching of *content* to the teaching of a *process* that allows students to access quicker mastery of similar content. Hattie (2009) organized Strategy Instruction in terms of the intended emphasis, including (a) learning intentions, (b) success criteria, (c) feedback, (d) student perspectives, and (e) student meta-cognitive/self-regulated learning. The strategies in all of these categories that rise to the top have in common the tendency to make instruction more explicit.

Among the strategies that emphasize learning intentions, Visual and Graphic Devices (.57) have the strongest effect size (Hattie, 2009). Visual and Graphic Devices, also called Concept Mapping or Non-Linguistic Representations (Marzano et al., 2001), is an instructional practice involving the creation of graphical or visual representations of lesson content with the intent of clarifying and synthesizing the big ideas with effect sizes recorded between .50 and 1.51. Visual and Graphic Devices make the relationship between concepts clear with explicit labeling and visual or pictorial representations of concepts or processes. For example, a Venn diagram can clarify the similarities and differences between two concepts using overlapping circles.

According to Hattie (2009), the most powerful strategies that emphasize success criteria include Mastery Learning (.58) and Worked Examples (.57). In Mastery Learning, the material is divided into small learning units that have clear, specific objectives and mini-assessments. Students must demonstrate minimal mastery in order to proceed to the next learning unit. Mastery Learning was one of the few specific strategies recommended by the NMAP (2006). Strong effect sizes are evident with students at all levels of schooling. Worked Examples is another example of a highly effective strategy, especially in math, that improves instructional clarity. The explicit goal of Worked Examples is to demonstrate to students what success looks like. Instruction consists of teacher demonstration of the steps required to successfully solve representative problems using similar examples and dissimilar or non-examples. Interestingly, Worked Examples instruction is made more explicit and effective when combined with Metacognitive Strategy instruction (ES=0.69)(Mevarech & Kramarski, 2003).

Feedback. Hattie (2009) determined that strategies that emphasize feedback are very effective. After synthesizing research on literally millions of students and thousands of teachers, he concluded, "I realized the most powerful single influence enhancing achievement is feedback" (p. 12). Great variability exists in the literature on the impact of feedback, however, indicating that more explicit or Elaborated Feedback is significantly more effective than Simple Feedback. Elaborated Feedback provides the student with a more detailed and explicit explanation of the correctness or incorrectness of the student response (Bangert-Downs et al., 1991). Corrective feedback is by definition more explicit and provides students with an explanation about what they are doing correctly and what they are doing incorrectly as they work towards a particular learning goal. Less explicit forms of feedback, such as praise or rewards are, unsurprisingly, the least effective, which Hattie (2009) attributes to the paucity of task related information provided.

Strategies that emphasize metacognitive strategies were all very powerful (ES=.69) according to Hattie's metaanalysis. Although complex, metacognition is defined most simply as "thinking about thinking." Metacognitive Strategies are activities that explicitly train students to be deliberate and thoughtful (i.e., clear and explicit) about their learning process in order to self-regulate their learning.

Marzano (1998) identified three key practices for teaching metacognitive skills: (a) providing students with specific learning objectives prior to every lesson (ES=.97), (b) providing feedback on the processes and strategies students use (ES=.74), and (c) giving students time to plan their approach to a task, then cueing specific thinking behaviors (ES=.53). Specific metacognitive strategies that repeatedly demonstrate impressive effect sizes include Student Think-Aloud (ES=.64) and Self-Evaluation (ES=.62). Especially effective in math, Student Think-Aloud involves the student's verbal expression of the internal problem solving process. Self-evaluation is defined as "setting standards and using them for self-judgment" and involves the student self-reflecting on their performance relative to their goals.

Frequent Formative Assessment. Some types of Formative Assessment have Very Large effect size, up to .90, for enhancing achievement. Formative Assessment can be any assessment of skills or understanding that is used explicitly to provide the student and teacher with feedback about what is learned and what is yet to be learned. It enables a teacher to modify instructional practices that are not improving student outcomes or continue the teacher practices that are. A standard definition of formative assessment is gathering information during instruction for purposes of gauging student learning. It answers the questions of "Where am I now?" and "What's next?" In contrast, summative assessment is the gathering of information after instruction for gauging what the student has learned.

Formative Assessment is often called Checking for Understanding and can vary on the continuum of explicitness depending upon the specificity of question/response during whole group instruction while circulating and listening as students practice new skills or work in pairs, by calling on random pairs to share their progress with the group, through quick-writes (e.g. students write a brief explanation or response to a prompt) or brief quizzes. A well-known and highly explicit form of Formative Assessment is weekly progress monitoring using tests like Curriculum Based Measurement (CBM) in reading or math (Fuchs & Fuchs, 1986; Fuchs, Fuchs, & Hamlett, 1994; Shinn, 1998). Weekly or daily quizzes that enable judgments of mastery can also be highly explicit. Effective Formative Assessment always informs instruction and often results in a teacher adjustment of instruction. Formative Assessment is also made more explicit and effective when the results are recorded or graphed by the student (i.e., supplemented with Self-Evaluation methods (Fuchs & Fuchs, 1986)). Formative assessment may be considered highly implicit when only brief unelaborated responses are required.

Teachers' Use of Effective Instructional Practices

Given poor mathematics outcomes nationally, perhaps it is not surprising that teachers are found to have limited knowledge of the instructional practices identified through generations of research on effective teaching. For example, in a study by Deshler et al. (2001) when 70 general education teachers asked to list five research-based teaching methods they use, they were often unable to do so.

Furthermore, many of the teacher-identified practices could not be termed "evidence-based". Of the instructional methods reviewed previously very few were consistently mentioned, only 8% of teachers listed direct/explicit instruction, and fewer than 3% listed graphic organizers and questioning techniques. Other evidence-based instructional practices listed by teachers were mentioned by fewer than 2% of respondents indicating a lack of knowledge of and general agreement about those most powerful evidence–based practices.

Observation studies typically confirm teachers' general lack of knowledge about effective instructional practices, and certainly their lack of application. For example, in an observation of 70 general education teachers at nine public high schools, Schumaker et al. (2002) found that teachers spent the vast majority of instructional time engaged in lecture or reading aloud to students, neither of which are considered powerful evidence-based practices by Hattie (2009). In some schools, these two activities accounted for an average of 94% of observed instructional intervals. Schumaker et al. concluded that teachers "engaged in few, if any, research-based instructional methods" for enhancing the learning of all students (p.13).

Similarly, in a recent study of high school effective teaching practices, Cornett (2010) found that few highly effective instructional practices were applied in high school classrooms. In fact, the most observed teacher behavior was disengagement from instructional activity (23.2%), (e.g., checking email or writing hall passes), followed by instructional activities such as Giving

Directions (21%) and Lecturing (12.8%). Meanwhile, evidence-based instructional activities found by Hattie (2009) to increase achievement were rarely demonstrated, if at all: Elaborated Feedback (8.3%), Modeling (1.6%), Graphic Organizers (0%), and Formative Assessment (0%).

Observational Studies of Effective Math Instructional Practices

Observational studies like Cornett's (2010) aimed at discovering how frequently research-based instructional practices are used are rare, and those targeting mathematics, and in particular algebra at the secondary level, are even more scant. The observational studies of secondary math practices are almost invariably focused on content, curriculum or teacher evaluation (Danielson, 2007; Pianta, La Paro, & Hamre, 2007) as opposed to pedagogy. Observational tools that are designed to focus on mathematics pedagogy are almost all designed based on NCTM standards to look for evidence of the use inquiry methods (e.g. student-led approaches), rather than evidence-based practices. The Reformed-Teacher Observation Protocol (R-TOP) (Sawada et al., 2000) and the Oregon Collaborative for Excellence in the Preparation of Teachers Classroom Observation Protocol (O-TOP) (Wainwright, Flick, & Morrell, 2003) are examples of such observation tools. One exception, the UTeach Observation Protocol (U-TOP) (Walkington et al., 2011) was purportedly designed to assess the overall quality of math instruction without preference or bias toward any particular way of teaching. However, the underlying dimensions of U-TOP were based on national reform standards such as the NCTM and NRC standards, and many of the items were highly inferential (Pianta et al., 2007). Observational studies focused on quantifying effective EB instruction specific to secondary Algebra classes were not to be found.

Purpose of the Study

The purpose of this study is to observe the instructional practices employed by mathematics teachers in order to determine the extent to which they are evidence based. Building upon studies like Cornett (2010), this study draws from converging research syntheses of highly effective practices in an attempt to answer the question of whether teachers use such practices in the context of Grade 9 Algebra 1 classes at large and ethnically diverse suburban high schools in Illinois.

CHAPTER 2-LITERATURE REVIEW

Increasingly, the national focus in education is on accountability and achievement outcomes. As described in Chapter 1, the majority of U.S. youth does not attain the necessary math and problem solving skills to participate, never mind contribute, to the global economy (or even to take care of their own fiscal futures). The pervasive U.S. student lack of algebra readiness by high school paired with the established importance of Algebra 1 creates a dilemma for American secondary educators. It is incumbent on Grade 9 Algebra 1 teachers to provide curriculum and instruction that addresses skill gaps and prepares students for future, even more challenging, course-work. The curriculum and instruction at the high school level must therefore be highly effective. Huge variance exists in U.S. math achievement and curriculum from one state, school, or one class to another (Kilpatrick et al., 2001; Loveless, Martin, Mullis, Schmidt, & Kilpatrick, 2008) suggesting that local differences in what and how we teach are extremely influential. Many students, therefore, do not receive consistent high quality instruction from one school to another, class to another, or one year to the next.

In response to the achievement problem, educators nationwide have focused on developing and increasing academic standards and rigor. Most recently, the National Governors Association Center for Best Practices & Council of Chief State School Officers (2010) published the Common Core State Standards in mathematics. The standards highlight expectations for what students should know when they complete secondary school to be prepared for college and career. The standards emphasize algebraic reasoning, especially the ability to make sense of quantities in problems and formulate equations to solve problems. Notwithstanding the importance of high and consistent standards, it is proposed here that too little emphasis is placed on ensuring that effective instruction is taking place in US Algebra 1 classes.

According to 40 years of research, three categories of instructional practices are consistently shown to enhance student achievement in math and across subject areas. These often overlapping methods, each of which can be said to improve teacher clarity, include explicit EB Strategy Instruction, Feedback, and Formative Assessment. It is the hypothesis of this study that student math achievement could be significantly improved if teachers recognized the importance of these foundational instructional practices.

Much of the data used to establish U.S. math underachievement is derived from international comparison data reported in Chapter 1. In this Literature Review, first briefly addresses some commonly held, but mistaken beliefs about the validity and usefulness of the international comparison data. Second, the quality of the research on highly effective teaching practices in the last 40 years is addressed, along with widely held concerns about the research to practice gap, the lack of knowledge transfer from educational research to classroom practice. Third, the contribution of the metaanalysis, the research method used for determining the comparative effectiveness of instructional practices included in this study, is elucidated in terms of its ability to recognize and synthesize patterns of evidence over many years and many findings in specific fields of study. Finally, the major aim of this literature review is to summarize the research supporting each of the most powerful evidence-based instructional practices that repeatedly rise to the top in education research. These highly effective instructional practices (IPs) fall into the three main categories of explicit Evidence-Based Strategy Instruction, Feedback, and Formative Assessment.

Is the International Comparison Data Valid?

In Chapter 1 the problem of U.S. math underachievement was established in large part by comparing the achievement outcomes of U.S. students compared to same age students in other countries. The validity of the international comparison data has been questioned over time with suggestions that maybe U.S. deficits have been overblown or that such comparisons are unfair. Do explanations promulgated in defense of U.S. performance on international tests hold up?

Explanation 1: U.S. student math performance at Grade 4 is strong. TIMSS data has shown that through Grade 4 American students are in fairly good shape by international comparisons of math ability, falling in the upper quartile of all countries (Mullis et. al, 2012). According to Hanushek et al. (2012), overall growth is

> disproportionately affected by 4th-grade performance, possibly leading to too much optimism. When we estimate gains only from student performance in 8th grade (on the grounds that 4th-grade gains are meaningless unless they are observed for the same cohort four years later) our results show annual gains in the United States of only 1% of a standard deviation (p.19).

Upon close analysis of international data Byron Auguste also confirmed, "the longer American children are in school, the worse they perform compared to their international peers" (McKinsey and Company, 2010). Sadly, the closer U.S. youth get to the job market, the less adequate are their skill sets.

Explanation 2: U.S. students are making gains. Indeed, U.S. students show some significant positive trends in achievement, however the much higher rate of growth in top performing countries (detailed in Chapter 1) all but renders these gains insignificant to close the gap. In fact, evidence suggests the gap is widening. "The measurable gain in achievement accomplished by more recent cohorts of students within the United States are being essentially matched by the measurable gains by students in the other 48 participating countries" (Hanushek et al., 2012, p. 12).

Explanation 3: The U.S. has many immigrant children to educate, compared to other nations. The U.S. was indeed ranked 6th among OECD countries in percentage of immigrant (19.5%) students. However, on closer inspection, PISA found only 3% of the variance among countries could be attributed to prevalence of students with immigrant status (Organisation for Economic Cooperation and Development, 2011).

Explanation 4: The U.S has a higher percentage of poor children to educate than other nations. In fact, the U.S. has about the same number of students considered economically disadvantaged as the average OECD country. However, PISA noted that 17% of the variance in learning performance could be attributed to socio-economic factors in the U.S. "Socio-economic disadvantage translates more directly into poor educational performance in the United States than is the case in many other countries" (OECD, 2011, p.2). In other words, poor and immigrant children are being more successfully educated in other countries. Furthermore, more than 80% of the variance in achievement cannot be explained by socioeconomic factors.

The U.S. keeps more children in school longer than other nations. Lastly, the international data has often been disputed based on the notion that in many countries lower performing students are weeded out of the system before high school, creating an unequal comparison, since U.S. students are required to attend school until the age of 17. While it is true the U.S. sets a relatively high compulsory education age among OECD nations, U.S. enrollment rates in primary and secondary education are the same as or below those in other OECD nations. Among OECD member nations, the U.S. ranks 23rd in enrollment of 15- to 19-year-olds. In fact, on the 2009 PISA assessment, OECD member nations on average tested a higher proportion of 15-year-olds than did the U.S. (89% versus 86% of the entire 15-year-old population) (Organisation for Economic Cooperation and Development, 2011). The idea that testing a broader population disadvantaged the U.S. may be considered a well-circulated myth (National Governors Association, Council of Chief State School Officers, & Achieve,Inc., 2008).

It is often the case that myth and popular or personal beliefs are so powerful that science or real outcomes are over shadowed or ignored. Germann (2010) stated, “beliefs are slow to respond to science and data”(p.4). The international comparison data reveals a serious problem for which educational researchers have uncovered some potential solutions if the science can make its way out of the laboratory. To do so, some mistaken beliefs about educational research may also need to be addressed.

The Rich and Underutilized Research Base

For the past 40 years, educational researchers have examined the vast database of research on general instructional practices to pinpoint those practices which deliver the best achievement outcomes for all students (Berliner, 1984; Brophy & Good, 1986; Cornett, 2010; Gage, 1984; Hattie, 2009; Marzano, 2007; Marzano et al., 2001; Rosenshine & Stevens, 1986; Walberg, 1984; Wittrock, 1984). The outcomes of these syntheses are similar in the clusters of teaching strategies that are repeatedly revealed to enhance student achievement, including explicit strategy instruction, feedback, formative assessment, and others.

Despite the depth and breadth of instructional strategy research, “some educators and non-educators hold a fairly low opinion of that research”

(Marzano et al., 2001, p.3). However, when educational research is compared to research done in other fields, such as medicine and physics, it stands up well to the scrutiny (Hedges, 1987). Berliner (1989) compares the attitudes towards research of medical professionals to that of educators.

> Indeed, many of the relationships between variables in educational research are stronger and more consistent than many of those on which medical practice is based. Yet few doubt the usefulness of medical research. Thus when we compare educational research with research done in other fields with established reputations for valid research, we find that educational research is of high quality. It follows that teacher educators have an obligation to take seriously the methods and findings of educational research (p.214).

Despite the high quality of educational research, comparable to quality found in hard science fields, it is often dismissed by educational practitioners. Educational researchers, in particular, lament that much quality research never makes it to the classroom. “We have a rich educational research base, but rarely is it used by teachers, and rarely does it lead to policy changes that effect the nature of teaching” (Hattie, 2010, p.2). The lack of transfer of pedagogical knowledge to practical use in the field of education creates frustration for some researchers. Lindsley (1992) publicly announced his refusal to spend a day more than the 25 years he had invested already in the field of educational research:

> Effective educational methods are available. They have been available for a long time. They are mostly behavioral, structured, fast paced, and require a high proportion of regular daily practice. Given this, it is irresponsible to invest more public funds on educational research without first installing the powerful results of the research we have already bought and paid for (p. 21).

The ultimate goal of all educational researchers is to see what is learned in the laboratory effectively translated to the classroom.

"Why does this bounty of research have so little impact?" (Hattie, 2009, p.3). One explanation provided by Hattie is that over-use and sometimes inappropriate use of the labels "evidence-based" and "research-based" muddies the decision-making waters for educators, which widens the research to practice gap. Deshler (2003) adds that curriculum publishers diminish the meaning of such labels:

> We need to be cognizant of the advertising potential currently associated with the phrase 'scientifically-based research.' Within the past several months, most publishers' booklists include several publications that claim to include "research-based" practices. While some of these publications are grounded in good research and careful trials across a variety of classroom contexts, it is clear that many are very loosely applying the term 'scientifically-based research'. Regrettably, this term can end up being used as political marketing and lose its intended meaning in the process (p. 6).

This loss of meaning is compounded according to Hattie (2009) by the plethora of evidence-based solutions that have come out of the education literature, resulting in the sense that "everything works."

Because navigating the educational research waters may be confounding and time-consuming for the average practitioner, it is incumbent on national, state, and district leadership to come to some agreement about what is good pedagogy based on solid research (Darling-Hammond, Wei, Andree, Richardson, & Orphanos, 2009). Fortunately, research methods such as meta-analyses makes it possible for researchers and practitioners to more efficiently identify results that are robust and replicated across time and settings.

The Meta-analysis

Gene Glass developed the idea of the metaanalysis in 1976. Prior to this time, research syntheses were more like integrated literature reviews (Hattie, 2009). Metaanalyses convert data from many studies meeting specific quality criteria in a particular area of research into an average effect size (ES). High quality metaanalyses utilize higher standards for study inclusion leaving out studies with significant validity concerns.

Effect size defined. Effect size is calculated by subtracting the average student score in a control group from the average score for students in a treated group, which is then divided by a measure of the variance or standard deviation in the sample. Effect size can be thought of as a standard deviation unit where 0 is average and an effect size of +1.0 is one standard deviation above average and an effect size of -1.0 is one standard deviation below average. Effect sizes close to 0 are interpreted as having little to no positive effect on the independent variable (e.g. achievement). In a meta-analysis, each study provides one or more effect sizes depending on the number of tools used to measure difference (between interventions) or change (over time). Average effect size is calculated by averaging the effect sizes across many different, but related studies. It is the average effect size of related studies that is used to determine the magnitude of effectiveness of an intervention or practice (Lloyd, Forness, & Kavale, 1988).

Determining strength of effect size. Prior to the advent of the metaanalysis, the emphasis was on statistical significance or the *p*-value. R. A. Fisher (1934) proposed the level $p = 0.05$, or a 1 in 20 probability of a result being exceeded by chance, as an acceptable limit of statistical significance, thereby suggesting the need for a difference of two standard deviations (on a normal distribution). Statistical significance, however, does not indicate the size or importance of an effect, leading to possible misunderstandings or misinterpretations of outcomes. Effect size does vary with statistical significance, for example, the larger the effect, the smaller sample size will be required to get a significant *p*-value. The use of effect size, especially in metaanalyses allows researchers to make statements about the potential magnitude of change over time or difference between interventions and can be translated into commonly understood metrics such as percentile gain. Lloyd et al. (1988) uses the example of an effect size of 0.67, which would indicate that the experimental group scores were on average higher than 75% of the control group scores. “In other words, if a student received the experimental treatment and had only an average score for all the students getting the experimental treatment, his or her score would be higher than the scores of 75% of the students who didn't get the experimental treatment”(p. 197).

Cohen (1988) suggested that an effect size of .8 to 1.0, one full standard deviation above average is *large*, therefore clearly seen or a blatantly obvious difference. Cohen further proposed that an effect size of .2 is *small*, and .4 is *medium*. Hattie (2009) in his metaanalysis of metaanalyses considered effect sizes of .2 *small*, .4 *medium*, and .6 *large* when assessing achievement outcomes.

Highly Effective Instructional Practices

Since the late 1970s, many meta-analytic studies and syntheses of the available instructional strategy research revealed that particular practices consistently proved to be highly effective. Some of the main past and present contributors to the research base on highly effective instruction include Gage (1984), Walberg (1984), Berliner (1986), Wittrock (1984), Marzano et al. (2001), and Hattie (2009). Summaries of their findings reveal undeniable patterns with regard to what works, with the oft-stated caveat that no one strategy works in every situation. Educators must therefore have a repertoire of evidence-based strategies at their disposable to differentiate instruction based on student needs.

As an example of an early synthesis, Walberg (1984) analyzed 170 books, 91 research syntheses on achievement, and surveyed 61 educational researchers. He determined three main causal influences on achievement, including student aptitude, instruction, and classroom environment. Among these influences on student achievement, he concluded that instructional quality variables, which are under the control of educators, were almost as influential as student aptitude, a presumably fixed variable. Under the classroom instruction category, Walberg discovered nine consistently effective methods: (a) graded homework (emphasis on individualized written and verbal feedback); (b) aligned time on task; (c) direct instruction, defined as systematic sequencing of lessons, guided student practice, feedback, and teacher clarity; (c) advance organizers connecting current learning to past learning; (d) teaching learning strategies; (e) tutoring; (f) mastery learning; (g) cooperative learning; and (h) adaptive learning, a system combining several of the first eight effective strategies. Walberg also determined that a variety of informal and formal goal-oriented assessments used to monitor progress and learning goals.

Another important early synthesis of highly effective instruction was contributed by Rosenshine and Stevens (1986) who determined that six "fundamental instructional 'functions'" were keys to improving student achievement (p. 379). The six functions included (a) review prerequisite knowledge, check and reteach previous day's work, (b) presentation of new content/skill (with clear modeling of procedures), (c) guided student practice (with checks for understanding), (d) provision of feedback and correctives (with re-teaching, if needed), (e) independent student practice with monitoring, and (f) weekly and monthly reviews (Rosenshine & Stevens, 1986). Studies conducted by Brophy and Good (1986) further established that these instructional components resulted in enhanced student achievement across grade levels and subject areas and for both basic and advanced skills. In addition, struggling learners benefit from the same clear teaching and instructional practices that benefit typical learners. (Baker, Gersten, and Lee, 2002, Gersten et al., 2006; Kroesbergen & Van Luitt, 2003).

Highly effective math instruction. In 1998, the National Research Council convened the Committee on Mathematics Learning in order to (a) synthesize the research literature on pre-kindergarten through eighth-grade mathematics learning, (b) provide research-based recommendations for teaching, teacher education, and curriculum and to identify future research needs, and (c) give guidance to educators, researchers, publishers, policy makers, and parents. With regard to highly effective mathematics instruction, the resulting document, *Adding It Up* (Kilpatrick et al., 2001) concluded that (a) a significant amount of class time should be spent in developing mathematical ideas and methods rather than only practicing skills, (b) questioning and discussion should elicit students' thinking and solution strategies and should build on them, leading to greater clarity and precision, (c) discourse should not be confined to answers only but should include discussion of connections to other problems, alternative representations and solution methods, the nature of justification and argumentation and (d) links among written and oral mathematical expressions, concrete problem settings, and students' solution methods should be continually and explicitly made during school mathematics instruction (pp.425-426). NRC's recommendations emphasized methods that improve teacher clarity through modeling and requiring mathematical thinking, explicit strategy instruction, problem-solving teaching, and frequent and ongoing formative assessment (i.e. checks for understanding) and feedback; consistent with non-subject specific highly effective instruction research.

Metaanalyses of algebra instruction. While the NRC's *Adding it Up* provided practitioners guidance regarding research-based math instruction, no such guidance existed specific to Algebra until 2006 when President George W. Bush created an independent review board, the National Mathematics Advisory Panel (NMAP). In the domain of instruction, NMAP asked what instructional practices are most effective for promoting algebra readiness (NMAP, 2008). NMAP acknowledged the complexity of teaching to groups with different backgrounds, interests, and motivation levels. NMAP determined that it is especially important for teachers to monitor student understanding and mathematical abilities, to use research when available to design appropriate instruction, and to purposefully utilize a mix of strategies matched to mathematical goals. A comprehensive teaching approach called Team Assisted Individualization (TAI) was found to be consistently effective in the teaching and learning of computational skills. The major components of this strategy included frequent formative assessment and mastery learning; assessment of target specific skill sets, skills are mastered in small groups by all students, and rewards are given based on evidence of skill mastery. NMAP attributed the success of this strategy to a higher frequency of feedback, the logical sequencing of skills learned to mastery, a motivating team approach with rewards, and the combination of teacher-directed and student-centered instruction. Like NRC, NMAP's recommendations were consistent with the previous universal findings suggesting that a relatively small number of instructional practices are repeatedly responsible for strong effects on student achievement.

For struggling math learners, which may be argued is a fair label for 65% of US students who, as shown earlier, do not meet minimum standards of proficiency on national tests, the (NMAP, 2008), recommended a few highly effective instructional practices. The NMAP suggested that low-achieving students do not differ significantly from identified math disabled students and therefore would benefit from similar instructional techniques that research has shown works with learning disability identified students: "Generally, clear consistent modeling of step-by-step strategies through teacher explanation, modeling and demonstration; careful control of task difficulty; planful sequencing of teaching and practice examples; and specified procedures for providing corrective feedback characterize explicit systematic instruction" (p. 48). The NMAP also suggested the use of concrete and visual representations, collaborative learning practices, and the meta-cognitive think-aloud strategy, requiring a student to verbalize their thinking. The What Works Clearinghouse (WWC) Response to Intervention (RtI) guide to effective mathematics instruction for struggling learners (R. Gersten, Chard, D.J., Jayanthi, M., Baker, S.k., Morphy, P., and Flojo, J., 2009) also recommends use of explicit teaching practices (i.e. modeling, guided practice, independent practice, feedback, and cumulative reviews).

Meta-analytic studies of algebra instruction. Very few meta-analytic studies have specifically focused on highly effective algebra instruction. However, an often-cited study by Marcucci (1980) marked a starting place for building a high quality math instruction knowledge base. From the 33 identified studies (11 focusing on Algebra) conducted between 1950 and 1980, Marcucci classified instructional practices into four categories, including Modeling (use of visual aids or manipulative materials) Systematic (a prescriptive approach to problem solving), Heuristic (the teaching of problem solving skills using diagrams or simplifying using smaller numbers), and Guided Discovery (questioning strategies employed to guide students discover solutions). Marcucci found that the Heuristic instructional method was the most effective for improving achievement in mathematics, but positive effects were significant only at the elementary level. At the secondary level (7-12), the Systematic approach, comparable to direct instruction methods, was the only grouping with significant positive effects.

Building on Marcucci (1980), Haas (2005) conducted a metaanalysis of 35 studies conducted from 1980 to 2002 at the secondary level where Algebra was the focus. The instructional practice categories were grouped on the basis of a literature review of algebra teaching methods, especially the studies included in the metaanalysis. Haas clarified that instructional practice groupings should not be considered mutually exclusive as "one teaching approach may contain another" (p. 27). For example, a direct instruction group may also have utilized technology, such as calculators for graphing. After two rounds of content validation, Haas identified six groups of instructional methods: Cooperative Learning (students working together towards common goal); Communication and Study Skills (teaching students to read and study math effectively and express math ideas verbally or in writing); Technology-Aided Instruction (use of calculator or computer-assisted instruction); Problem-Based Learning (teaching deductive and inductive reasoning for problem solving); Manipulatives, Models, and Multiple Representations (teaching concepts using concrete or symbolic representations); and Direct Instruction (DI) (explicit review of previously learned content and the sequential teaching of steps, paired with practice and feedback). DI had the highest effect size (ES=.55) regardless of student ability. Problem-Based Learning (PBL) also showed a high effect size (ES=.52), but had medium effects with the lower ability classes studied with a small negative effect on higher ability students. Manipulatives, Models and Multiple Representations had a medium effect size (.38).

Rakes et al. (2010) conducted the most recent metaanalysis to determine what are the most effective instructional methods used in teaching Algebra at the secondary level and furthermore what are the characteristics of Algebra instruction that make the biggest difference in achievement. Rakes et al. included studies from 1968 to 2008, searched 20 electronic databases and out of 594 potentially relevant studies retained 82 studies that met strict inclusion criteria. Instructional Strategies (ES=.35) and Manipulatives (ES=.34) were two of the five categories identified as significantly enhancing achievement, along with Technology Tools (ES=.17), Technology-Based Curriculum (ES=.15), and Non-Technology Curriculum (ES=.40). Instructional Strategies was made up of four individual strategies including Cooperative Learning, Mastery Learning, Multiple Representations (i.e. Visual and Graphic Devices), and Formative Assessment strategies. Manipulatives was defined as the use of concrete objects to enhance understanding, such as the use of tiles or geometric cutouts. Although each meta-analysis grouped variables slightly differently, strategy instruction consisting of DI, Mastery Learning, use of Visual and Graphic Devices, Feedback, and strategic Formative Assessment practices consistently demonstrate the highest effect sizes in the professional literature focused on secondary algebra and general mathematics instruction.

Putting it all together. Early research syntheses, the recommendations of national panels of experts, and Algebra-specific meta-analyses are corroborated by recent large-scale meta-analytic studies conducted by Marzano et al. (2001) and Hattie (2009). Marzano et al. identified nine high effect-size instructional practices to assist teachers in becoming highly effective. These instructional practices were Identifying Similarities and Differences, Summarizing and Note-taking, Reinforcing Effort and Providing Recognition, Homework and Practice, Non-linguistic Representations, Cooperative Learning, Setting Objectives and Providing Feedback, Generating and Testing Hypotheses, and Cues, Questions, and Advance Organizers. See Table 2 for a listing of each IP and its associated definition and average effect size.

Table 2

Marzano et al. (2001) categories of high effect size instructional strategies.

Instructional Practice Category	Definition	Ave. Effect Size (ES)
Identifying similarities and differences	Teaching of the similarities and differences between topics or problem types often involving explicit modeling and graphic representations.	1.61
Summarizing and note-taking	Explicit teaching of the skills needed to synthesize learned content especially through the use of summary framing (series of structured teacher-led questioning).	1.00
Reinforcing effort and providing recognition	Explicit teaching of the importance of believing in effort.	.80
Homework and practice	Time allowed for practice with goal of skill mastery, starting with a guided shaping phase until ready for independent practice geared towards accuracy and speed.	.77
Nonlinguistic representations	Explicit teaching using graphic, visual, or physical representations of concepts.	.75
Cooperative learning	A grouping strategy involving student collaboration, interdependence, positive peer support, group accountability, skill mastery, and group processing.	.73
Setting objectives and providing feedback	Goal setting and corrective, specific, and timely feedback that is criterion or skill related versus norm-referenced.	.61
Generating and testing hypotheses	Teaching deductive reasoning whereby students use a general rule to understand similar problems or predict future actions or events. Inductive reasoning techniques show lower effect sizes (ES=.39)	.61
Questions, cues, and advance organizers	Strategies used by a teacher to activate prior knowledge.	.59

*Note***.** Categories and definitions adapted from “Classroom Instruction That Works” by Marzano, Pickering, and Pollack (2001).

In a similar vein, Hattie (2009) synthesized over 800 meta-analyses relating to student achievement. Hattie credited the best effects to be occurring in purposeful interactions between teachers and students through strategy instruction and a feedback and assessment loop. Hattie indicated that the effects of what he called “visible learning” are strong for both low skill learners and typical learners across subject areas. “The teacher needs to invite the student to learn, provide much deliberative practice and modeling, and provide appropriate feedback and …independent practice” (p.207). Among Hattie’s top performing instructional practices were Formative Evaluation, Feedback, Spaced Practice, Metacognitive Strategies, Problem-Solving Teaching, Teaching Strategies, Direct Instruction, Mastery Learning, Worked Examples, and Visual and Graphic Organizers. Each of these evidence-based instructional practices is labeled and defined in Table 3 with a summary of the underlying research provided in the next section.

Table 3

Hattie (2009) categories of high effect size instructional strategies.

Instructional Practice Category	Definition	Ave. Effect Size (ES)
Formative Evaluation	Assessments that are used by teachers as feedback to adapt instruction to students' needs, even more effective when student data is graphed.	.90
Feedback	The most effective feedback is related to learning goals and is timely, allowing for immediate student action.	.73
Spaced Practice	Deliberative practice over days considering number of correct and incorrect student responses with gradual release to independent practice with a goal of mastery and fluency.	.71
Metacognitive Strategies	Explicit teaching of self-monitoring and problem-solving involving "thinking about thinking." Assists students in becoming self-regulated learners.	.69
Problem-Solving Teaching	Teaching deductive reasoning or a heuristic to understand similar problems.	.61
Direct Instruction	An explicit teaching approach based on a sequence of instructional practices including identifying learning goal, activation of prior knowledge, explicit modeling, guided practice, independent practice and assessment.	.59
Mastery Learning	Teacher provides clear success criteria and sequences small learning units that must be mastered before new material is presented.	.58
Worked Examples	Teacher provides students with a completed problem example along with the steps towards the solution.	.57
Concept Mapping	The use or development of visual or graphic devices to understand concepts	.57

*Note***.** Categories and definitions adapted from "Visible Learning" by Hattie (2009).

Themes of Evidence-Based Instruction

The professional literature reveals that researchers often differ in the way individual instructional practices are labeled and grouped. The variability is explained well by Haas (2005) who emphasized the point that the study of any individual strategy in real time does not preclude the simultaneous use of other effective strategies. For the purposes of this study highly effective instructional practices were grouped into three primary categories, but it should be noted that highly effective instruction almost always includes a combination of all three components. The three evidence-based categories labels are: a) Evidence-Based Strategy Instruction, b) Feedback, and c) Formative Assessment. Each instructional practice category includes discrete highly effective instructional practices with average effect sizes greater than .55 according to the Hattie (2009) or Marzano et al. (2001) metaanalyses. The discrete practices chosen for review consistently demonstrate *Medium* (ES=.55-.60) to *Large* (ES >.60) effects on achievement.

A theme discovered across all of the highly effective IPs is a high level of explicitness. Strategy instruction can vary from highly implicit, as with inductive techniques (e.g. PBL or other constructivist approaches) where the student must discover principles from real-world problems, to highly explicit form in the sequential steps of Direct Instruction (DI). Archer and Hughes (2011) argued that "in some ways instructional approaches can be put on a continuum of how much guidance and scaffolding are considered desirable in teaching new skills to novice learners or intermediate learners” (p.18). A well-circulated belief is that “drill and practice” (associated with DI) interferes with student creativity or does not lead to the development of higher order thinking (Heward, 2003), but no research supports this claim. Others are concerned that isolated skill instruction may result in the students’ inability to generalize learning to an overall skill set (Poplin, 1988; Stainback & Stainback, 1992). Long-term and working memory research indicates that expert problem solvers derive their skills and abilities from their long-term memory of a topic. Because they know a lot about a topic they are able to construct meaning or discover solutions with minimal guidance (Kirschner, Sweller, & Clark, 2006). Conversely, cognitive overload or the over-taxing of the working memory occurs in novice and intermediate learners, which interferes with the ability to focus on key concepts. A high level of guidance, i.e. clear models, demonstrations, explicit review, and deliberative practice, reduce the load on working memory and is shown to be highly effective with novices and as effective as alternate methods for students with prior knowledge (Kirschner et al., 2006).

Evidence-Based Strategy Instruction

The first category of evidence-based instruction, EB Strategy Instruction, includes the subgroups of DI, Metacognitive Strategies, and Other Strategies.

Direct instruction. DI programs are characterized by a focus on skill mastery via a sequence of seven explicit instructional practices: goal clarification, explanation of success criteria, explicit modeling, checking for understanding, guided practice, independent practice, and explicit review. In DI, explicitness is ensured by the use of teacher modeling of the procedures involved in learning a strategy, especially with new material. Sufficient guided practice followed by independent practice of each new skill is provided and strategies and skills are explicitly reviewed to ensure they are learned and generalized within days and weeks of the new learning.

Direct instruction methods have been shown to be highly effective with elementary and high school students in general education (ES=.99) and special education (ES=.86) (Adams & Englemann, 1996). Most recently, Hattie (2009) found DI methods across four metaanalyses, 304 studies, and 42,000 students, had an effect size of .59. Przychodzin-Havis, Marchand-Martella, Martella, and Azim (2004) found that out of 12 DI studies specifically applied to mathematics 11 showed significant positive effects. Schmoker (2011) emphasizes that the essential parts of a good lesson include a clear learning objective with review of pre-requisite knowledge, teacher-led modeling, guided practice, checks for understanding, and independent practice; “the most simple, ordinary teaching strategies overcome all other factors by significant margins” (p. 12).

Explicit Modeling. Explicit Modeling is a key component of DI programs and highly effective strategy instruction. Explicit Modeling consists of a demonstration and description of a skill. Modeling is defined as a teacher think-aloud giving students access to the thinking process of the teacher, including the types of decision-making used while solving a problem. A good model is clear, concise, consistent, includes several demonstrations and involves the students (Archer and Hughes, 2011). Highly effective teacher-led Explicit Modeling goes through the procedural steps of problem solving and also clearly demonstrates the needed scaffolding for students to understand simple to increasingly complex concepts. Cornett (2010) emphasized the importance of distinguishing between explicit and implicit modeling; explicit modeling indicates a physical demonstration of the steps with verbalization of the teacher's thought process and implicit modeling indicates a demonstration without the verbalization of the thought process (e.g. silently working out a problem on the board). Archer and Hughes (2011) defined Explicit Modeling as the "I do it" phase of learning occurring when teaching new skills or content when the teacher orally walks students through the steps of a skill, concept, or strategy. The Explicit Modeling of examples and non-examples is often recommended (Archer & Hughes, 2011). Similarly, Marzano et al. (2001) cites a trove of research indicating strong positive effects on achievement with the explicit and strategic presentation of similarities and differences.

Explicit Review. Explicit Review, another important component of DI, is the practice of revisiting procedures and concepts previously taught. Research indicates that appropriate review may affect the quality of what is learned both in terms of recall and concept development and enhances a student's ability to solve complex problems (Dempster, 1991). The explicit review of topic vocabulary or essential terms is especially key for improving student comprehension (Marzano, 2012) and to avoid language problems that often underlie mathematical misunderstandings (Kilpatrick et al., 2001). Effective review is necessarily guided by information gleaned from informal and formal assessments that require the use of the prerequisite skill (Cornett, 2010) and is used to verify that students have mastered pre-requisite skills prior to teaching new concepts (Archer & Hughes, 2011). A review of success on homework geared to prerequisite skills can serve this function as well as guided practice with feedback (Archer & Hughes, 2011; Marzano, 2007). Research consistently indicates that the use of reviewing and summarizing strategies significantly improves student achievement (Hattie, Biggs, & Purdie, 1996; Marzano et al., 2001; Rosenshine & Meister, 1994).

Guided Practice. Practice is an essential component of explicit teaching with the ultimate goal of providing a gradual release of responsibility from the teacher to the student while learning to mastery. Guided Practice is often called deliberative practice as it is skill focused and well-planned, rather than haphazard. Archer and Hughes (2011) defined Guided Practice as the “We Do” phase of explicit instruction and Independent Practice as “You Do.”. Archer and Hughes suggest that Guided Practice is provided through targeted verbal and visual prompts, directions, clues, cues, or reminders provided while the student performs the new skill. Good, Grouws, and Ebmeier (1983) noted that with new material, guided practice contributes to higher levels of success on later independent work. In addition, planned prompt timing during guided practice leads to enhanced achievement and reduced error rate (Cybriwsky & Schuster, 1990). During Guided Practice, the learning needs of individuals and the group become clear and teachers are able to scaffold, through feedback, informal assessment and instructional adjustments.

> In classes with diverse learners, it is virtually certain that some students will have difficulty even with the best crafted explanations…Students are not apt to be successful doing independent work unless they first demonstrate high levels of success in the controlled practice phase of instruction (Carnine, Jones, & Dixon, 1994, p. 421).

Independent Practice. Research in cognitive psychology suggests that learning progresses in a specific way and much more practice is needed for students to master a new skill than is typically provided by U.S. teachers (Healy, 1990). For example, studies have shown that students do not reach 80% accuracy on average until after practicing skills as many as 24 times (J. R. Anderson, 1995). Independent Practice follows deliberative Guided Practice to increase skill fluency or automaticity, but only when students' error rates are low enough that they will not practice skills incorrectly. According to research, "the amount of time that students are actively engaged in learning skills at a high rate of success is positively related to the acquisition of those skills" (Carnine et al., 1994, p. 422). Spaced versus massed practice is found to have the most impact on achievement, both for acquiring and retaining new skills or information (Donovan & Radosevich, 1999). Less space between practice sessions is needed for surface or simple skills, while longer spacing is more effective for more complex tasks. Hattie (2009) conjectured that the common denominator of many of the most effective teaching methods, such as direct instruction, peer tutoring, mastery learning, and feedback is deliberative practice.

Metacognitive Strategy Instruction. Metacognition, or thinking about thinking, involves self- questioning and self-regulatory behaviors that direct a person to understand how their thoughts and actions relate to successful problem solving. Strategies that emphasize metacognitive strategies were very effective (ES=.69) according to Hattie's (2009) meta-analysis. Metacognition is activated when individuals are faced with challenging problems (Montague & Applegate, 1993) and is integral to effective mathematical problem solving (Garofalo & Lester, 1985; Schoenfeld, 1985, 1987, 1989, 1992). Students who have better metacognitive abilities perform better in mathematical problem solving (Artzt & Armour-Thomas, 1992; Carpenter & Fennema, 1996). Unfortunately, research indicates that students often do not recognize or select appropriate strategies when solving mathematical problems (Schoenfeld, 1985). Fortuitously, studies also suggest that metacognitive skills can be learned (Garofalo & Lester, 1985; Montague, Enders, & Dietz, 2011). Furthermore, according to Montague et al. (2011), the explicit teaching of metacognitive strategies such as the *Solve It!* strategy has been shown to equally benefit students with an identified mathematical learning disability, low achieving math students and average achieving math students. The Zimmerman (2000) cyclical model of self-regulation suggests that self-regulated learning is defined by a cycle of three sequential phases, (a) forethought, (b) performance, and (c) self-reflection. The forethought phase includes task analysis and sources of motivation, such as self-efficacy beliefs. The performance phase involves self-control and self-observation; use of various strategies to complete a task and monitoring of one's performance, such as self-questioning. The self-reflection phase occurs after engaging in a task and is involving self-judgments or self-evaluation.

Problem Solving Teaching. Problem-Solving Teaching, a strategy aimed at the forethought phase of metacognition, was also identified by Hattie (2009) as one of the top highly effective instructional practices (Average ES=.61), also called the heuristic method. Mathematics problem solving can be understood using Polya's (1957) highly influential four-stage model, which suggests that successful problem solvers, a) understand the problem, b) devise a plan, c) carry out the plan, and d) check that the solution was reasonable. Problem solving therefore requires the ability to remember and execute procedures, but also to choose appropriate methods, plan and self-monitor, a combination of cognitive and metacognitive processes (Goldman, 1989). Teachers trained in heuristic methods of problem solving, such as mnemonic devices, realize better student achievement outcomes (Hattie, 2009; Marcucci, 1980). Hattie et al. (1996) found that the explicit teaching of mnemonics or problem-solving steps, such as Polya's four-stage model, was highly effective across age groups and subject areas (ES=.79). Marzano et al. (2001) also identified problem solving teaching, especially deductive reasoning techniques, as highly effective (ES=.61). Deductive problem solving implies that the rules or principals underlying a problem solution are provided to the student, rather than discovered, as with inductive techniques.

Student Think-Aloud. Student Think-Aloud, also called self-verbalization/self-questioning, is an example of a highly effective meta-cognitive strategy aimed at the performance phase with an effect size of .64 (Hattie, 2009). Especially effective in mathematics (Duzinski, 1987), Student Think-Aloud involves the student's verbal expression of the internal problem solving process, useful both for enhancing retention of successful processes, but also revealing errors in thinking to the student and the teacher, allowing for corrective feedback. Ostad and Sorensen (2007) found that task relevant think-aloud was associated with successful task completion. Hattie (2009) attributes the success of such strategy instruction to students becoming active in the learning process, in effect learning how to learn. Student Think-Aloud, likewise, forces the student to identify what they understand and to clarify what information they need.

Self-Evaluation. Another highly effective Metacognitive Strategy, Self-Evaluation, is defined by as "setting standards and using them for self-judgment" (Hattie, 2009, p.190) and involves the reflection phase of metacognition; the student reflects on their performance relative to their goals. Self-evaluation (ES=.62) has a higher effect size than simple self-monitoring (ES=.45) as it requires more active involvement of the learner and clarity about their progress (Hattie, 2009). The Self-Evaluation strategy has been found to be most effective when student's self-judgments are relatively accurate. "Students will only be able to self-regulate their learning effectively if they monitor and evaluate their progress accurately and thus make adaptations that are based on a correct analysis of their performance"(Labuhn, Zimmerman, & Hasselhorn, 2010). Unfortunately, many empirical studies of student self-evaluation accuracy reveal that most students lack skills to estimate their performance accurately (Labuhn et al., 2010). The Self-Evaluation strategy is not different than most effective strategies and can be improved with explicit instruction and feedback. Feedback is an inherent component of self-regulated learning providing learners with information about how well they are performing and increases their awareness of the quality of their achievement and motivates future self-monitoring behavior (Butler & Winne, 1995).

Reinforcement of Effort or Persistence. Students' self-assessments are also impacted by their beliefs about the path to success. When students do not believe they have ability to succeed they may self-sabotage (Marzano et al., 2001). Further, students do best when they attribute success they do have to effort, rather than to another person, ability, or luck. The importance of Reinforcing Effort/Persistence is demonstrated in numerous studies cited by Marzano (e.g.,Seligman, 1990; Urdan, Midgely, & Anderman, 1998), wherein many students were found to not be aware of the relationship between their own effort and their potential success on a task. Explicit teaching of Reinforcing Effort/Persistence has been demonstrated to have consistently *Large* effect sizes (Average ES=0.52 to 2.14) and this type of feedback can have a larger impact on student achievement than time management and comprehension techniques (Marzano et al., 2001).

Metacognition is triggered when students are faced with challenging problems and importantly, the students' perception of the difficulty of a problem has an effect on their persistence in solving the problem (Montague & Applegate, 1993). Sfard (1991) found that both students and teachers often expect immediate rewards for teaching and learning efforts, but "the reification, which brings relational understanding, is difficult to achieve, it requires much effort, and it may come when least expected, sometimes in a sudden flash"(Sfard, 1991, p. 33). Several techniques proposed for helping students make the connection between their effort and their ability to succeed, including explicitly teaching and emphasizing the connection through real life stories or by using a self-assessment rubric after completion of tasks in class and even graphing their results over time. This type of feedback is perhaps invaluable as it is generated by the student and supports the crucial development of self-regulation skills.

Other Strategy Instruction. Several other strategies consistently obtain high effect sizes for student achievement, including the use of Visual and Graphic Devices, Worked Examples and Mastery Learning.

Visual and Graphic Device. A Visual and Graphic Device is a visual representation of an idea or a concept used to organize and remember new information assisting students in making connections with previously learned content (Alvermann, 1981; Robinson & Kiewra, 1995). Both Marzano et al. (2001) and Hattie (2009) suggested that Visual and Graphic Devices are more effective when used to organize and synthesize information. Visual and Graphic Devices highlight the important aspects of a skill or topic and show in words and pictures the relationships between concepts previously and currently being taught. Visual organizers provided at the beginning of a lesson have been shown to increase the retention of new material (Lenz, Alley, & Schumaker, 1987). Visual and Graphic Devices can be used with any subject matter and at almost any grade level to support a wide variety of learners (Bromley, Irwin DeVitis, & Modlo, 1999). The conceptual roots of Visual and Graphic Organizers are in schema theory, which states that new information must be linked to pre-existing knowledge (Ausubel, 1968). Used at the beginning of a lesson graphic organizers help students to activate prior knowledge and to connect the new information to the old, enhancing both understanding and retention. Merkley and Jefferies (2001) formulated guidelines for using graphic organizers more effectively than simply using them as a form of lecture notes without student participation. They recommended that teachers a) verbalize the relationships or links among concepts expressed by the visual, b) allow students to interact or discuss the graphic, c) explicitly connect the new learning to past learning, and d) make predictions about what learning comes next.

Worked Examples. The Worked Example is another highly effective and highly explicit instructional strategy (Hattie, 2009). A Worked Example is a written problem that is solved with all of the steps included and provided to the students. The Worked Examples strategy is based on schema theory; schemas are "mental constructs that allow patterns or configurations to be recognized as belonging to a previously learned category and which specify what moves are appropriate for that category" (Sweller & Cooper, 1985, p. 60). Studies show that successful problem solvers use previously acquired schemas and novices use general search strategies often leading to the use of inefficient schemas (Cooper & Sweller, 1987; Lange, Booth, & Newton, 2014; Sweller & Cooper, 1985). Worked Examples prevent novice students from using such inefficient strategies leading to the development of proficiency. In laboratory studies on the effect of worked examples on learning Algebra and other high school mathematics, students who were presented with multiple worked examples learned as well or better than did students who were presented with a few examples followed by conventional practice, even with half the allocated practice time (Cooper & Sweller, 1987; Lange et al., 2014; Sweller & Cooper, 1985). Another study found that "students who were given Worked Examples required "less acquisition time, needed less direct instruction, made fewer errors, and made fewer types of errors during practice …and learning transferred to test even in the absence of the worked example" (Carroll, 1992, p. 365; 1994). In the same study, on a delayed post-test, the Worked Example group made a quarter of the errors of the conventional practice group and low achieving students made significant gains towards closing the gap with average achieving peers. Zhu and Simon (1987) reported that students exposed to Worked Examples in Algebra mastered topics in less time than in a typical classroom situation and that the students' learning went beyond memorization towards considerable depth of understanding.

Many recent studies of Worked Examples include the use of Worked *Non*-Examples. Worked Non-Examples are problems that are solved incorrectly or inefficiently. When students are asked to explain incorrect procedures they begin to understand how and why the incorrect procedure is wrong. Siegler and Chen (2008) found that explanations of why incorrect examples are incorrect are more effective than explanations of why correct examples are correct.

> The benefits of explaining errors draws students attention to the particular features of the problem that make the procedure inappropriate, helps students replace faulty conceptual knowledge they have about the meaning of the problem feature with correct conceptual knowledge about those features (Booth, Lange, Koedinger, & Newton, 2013, p. 5).

Other studies found that purposefully timing the introduction of non-examples may be particularly important as some prior knowledge appears to be needed to fully benefit from the strategy. When novice learners are expected to locate the error in non-examples themselves, results have not led to increased benefit (Grobe & Renkl, 2007). Worked Examples and Non-Examples are highly explicit strategies that have been shown to enhance both procedural and conceptual knowledge in mathematics.

Mastery Learning. Mastery Learning is an instructional practice based on clarity of expected learning outcomes. Mastery goals are generally divided into small sequential learning units or discrete skills, each of which is followed up by formative assessments used both as evidence of minimum proficiency and to identify learning gaps to guide further instruction. Hattie (2009) found that Mastery Learning had an average effect size of .58 for achievement, but also strong affective outcomes (e.g. motivation, engagement). Kulik, Kulik, and Bangert-Down's (1990) meta-analysis of 103 studies primarily focused on Mastery Learning in the upper grades showed an effect size of .52 with the average treatment group score at the 70th percentile and the average control group score at the 50th percentile. Additionally, gains for lower skilled students (ES=.61) were more positive on average than gains for average students (ES=.40), but the difference was not statistically significant and *medium* to *large* gains were made by both groups. When pursuing mastery goals, students focus on developing skills and success is based on personal progress rather than competition with others (Senko & Hulleman, 2013). Mastery Learning often involves a great deal of cooperation between students and high levels of feedback and formative assessment to ensure students are progressing and continue to be challenged (Hattie, 2009).

Feedback

Feedback can reinforce effort, provide recognition, or provide correction. Feedback, as defined by Hattie, is "a consequence of performance" (p.174), it is information provided that allows the learner to "confirm, add to, overwrite, tune, or restructure information in memory" (Winne & Butler, 1994, p.5740). Hattie (2009) found that across 23 meta-analyses of Feedback interventions, the average effect size was Large (ES=.73). However, Hattie qualifies these positive effects due to great variability that exists in the literature on the impact of feedback, indicating that not all feedback enhances learning equally.

Three kinds of feedback are found to enhance achievement, including Task Level (e.g. How well tasks are understood/performed), Process Level (e.g. the process needed to understand/perform tasks), and Self-Regulation Level (e.g. self-monitoring, directing, and regulating of actions) (Hattie, 2009). A fourth type of feedback, Self Level, (e.g. personal evaluation of the learner) is considered ineffective and has even been shown to lead to decreased effort, even when positive, as it draws attention to the self over the task (Black & Wiliam, 1998). Feedback is understood to be most effective when it is task or goal-related. Based on a meta-analysis of Feedback interventions, Kluger and DeNisi (1998) suggest that Feedback that is directly related to set goals is significantly more effective than general Feedback. Marzano et al. (2001) combined Setting Objectives and Providing Feedback for one of his top most effective instructional practices with a *Large* average effect size of .80. Marzano et al. also differentiated between simple feedback (providing information that solution is right or wrong only), a *small* effect size practice (ES=.22) and corrective and criterion-referenced (task-specific) feedback that consists of more elaborated explanations (ES>.90).

Teachers have been found to overestimate the frequency that they provided detailed task-oriented feedback (Nuthall, 2005). Of further concern, low ability students, students from backgrounds of poverty, and students of color have been shown to receive less feedback, are given less time to respond, asked less demanding questions, are criticized more, and are called on less (Ladson-Billings, 1997). Research also shows that more frequent feedback is especially beneficial for students with less prior knowledge (Fyfe, Rittle-Johnson, & DeCaro, 2012).

Formative Assessment

Feedback and formative assessment could be seen as two sides of the interactive teaching and learning coin, inextricably linked. Effective formative assessment triggers feedback and vice versa. Regarding Formative Assessment, Black and Wiliam (2009) explained:

> Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited (p. 9).

Black and Wiliam further explained that Formative Assessment is (a) about "moments of contingency" (p.10); (b) for the purpose of the regulation of the learning process; and (c) frequently precedes real-time adjustments in whole group or individual instruction. Formative Assessment can be accomplished through question/response during whole group instruction, by circulating and listening as students practice new skills or work in pairs, by calling on random pairs to share their progress with the group, through quick-writes (e.g. students write a brief explanation or response to a prompt) or brief quizzes or tests, but it is the instructional decision made after data is collected that makes assessment formative. Marzano (2007) advocated "continually checking for understanding" (p.87) and ensuring that every student is responding multiple times while learning new material. Although frequent checks for understanding is known to slow down instruction, research has shown that doing so can dramatically increase student understanding and can account for an additional six to nine months of learning per year (Black & Wiliam, 1998; Marzano, 2007). Among studies that show the impact of formative assessment on achievement, the most effective have tended to be brief often informal assessments that impacted teachers' day-to-day and minute-to-minute classroom practices (Leahy, Lyon, Thompson, & Wiliam, 2005).

Progress monitoring is a more systematized method of formative assessment involving short paper-pencil tasks assessing discrete and often sequential skills. Hattie's (2009) metaanalysis of meta-analyses found that providing formative progress monitoring had the highest average effect size of all instructional practices in the teacher's domain (ES=.90). Fuchs and Fuchs' (1986) meta-analysis of effects of frequent progress-monitoring and subsequent adaptation to instruction indicated a .70 effect size for improved student performance when the teacher participated in this type of ongoing data review combined with adjusting instructional practices as a response to the information collected through assessment.

Summary

According to 40 years of research, three categories of instructional practices are consistently shown to enhance student achievement in math and across subject areas. These often overlapping methods, each of which can be said to improve teacher clarity, include explicit strategy instruction, feedback, and formative assessment. Given the research base has both depth and breadth and research methods are very often up to the highest standards of science, it stands to wonder to what degree the instruction in real-life classrooms are influenced. It is the hypothesis of this study that many evidence-based practices are under-utilized and student math achievement could be significantly upgraded if teachers are better exposed to and trained in these foundational instructional practices. Specifically, this study seeks answers to the following research questions:

1. What instructional practices do Grade 9 Algebra 1 teachers employ when teaching?
2. How often are evidence based instructional practices used, with respect to estimated frequency and duration?
3. Are there differences in observed instructional practices as a function of student achievement level?

CHAPTER 3- METHODS

The purpose of this study was to observe systematically the instructional methods that Grade 9 high school algebra teachers use and the degree to which these methods include the evidence-based practices. This study will employ systematic observation with an instrument and strategies adapted from the study of evidence-based instruction by Cornett (2010).

A Locally Representative Sample

The participating high schools, based on demographic, socio-economic, and test scores look like average high schools in Illinois and Illinois is in the middle when compared to the national math achievement statistics. Despite the fact that the current sample was chosen based on convenience, e.g., the researcher worked in the participating school district in one of the participating high schools, these surface similarities may contribute to improved generalizability of findings.

Grade 9 algebra classrooms from two high schools from a public consolidated school district from the northwest Chicago suburbs served as the participants for this study. The school district's total 2012-13, K-12 enrollment was 20,266 students with 5,977 of them attending three high schools. Descriptive information on the participating schools is presented in Table 4 where data on racial and socio-economic factors, dropout and truancy, and average ACT scores are presented. For the purposes of the study, one school will be Diverse High School and the other Suburban High School.

Table 4

2013 Demographic information for the two participating high schools.

School/District Characteristics	Diverse High	Suburban High
Total Enrollment	2,586	2,186
White	39.6	73.8
Black	7.8	2.8
Hispanic	47.5	14.6
Asian	2.2	6.3
Mixed Race	2.6	2.3
Low-Income	61.8	20.6
ELL	7.6	1.1
Graduation Rate	82.0	93.0
Chronic Truancy Rate	25.6	5.7
Meets or Exceeds on PSAE Math	42.0	52.0

Note. Other than enrollment, numbers represent percentage of total enrollment. PSAE refers to the Prairie State Achievement Exam.

Setting. As displayed in Table 4, Diverse High is the larger and more racially diverse of the two participating high schools; it also serves a high percentage of students (62%) from low-income families. The dropout rate and chronic truancy rate of Diverse High is also elevated compared to Suburban. Diverse High is located in a suburban area, with many economically depressed neighborhoods and some affluent areas.

Suburban High is located in a middle-class suburban environment. Suburban High is less racially and economically diverse than Diverse High serving a predominantly White student population and only 1 in 5 students come from low-income families.

Classroom Participants. Two types of Grade 9 algebra classes are taught at each of the participating high schools, each with a different student achievement-level focus and a different amount of allocated time, a one-period (1P) and a two-period (2P) class. The 1P and 2P algebra classes use the district-approved curriculum guided by the Illinois Edition of the Glencoe Mathematics Algebra 1 textbook (Holliday et al., 2005) and share standards-based common assessments.

A general education algebra instructor taught each 1P and 2P class. Some of the classes were "co-taught," where two teachers, a general educator and a special educator, shared instructional responsibilities. The special educator's role differed class-to-class, varying from a shared role where both lead whole group discussions to one in which the special education teacher assisted only struggling students with a special education label. The instructional activities of both the general educator and special educator were observed and recorded in all co-taught classrooms.

2P and 1P classes also occasionally employed a paraprofessional whose role varied from circulating the room and monitoring student behavior to assisting individuals or groups of students. Although the presence of a paraprofessional in classrooms was noted, the instructional activities of the paraprofessionals were not recorded for this study.

1P Algebra. The 1P Algebra I class is scheduled for 45 minutes per day and is designed for students who have average to above average math skills according to their Grade 8 Explore test (ACT, 2013). Explore is a test that provides information about how well a student performs compared to other students with regard to ACT's college readiness standards (ACT, 2013).

2P Algebra. The 2P Algebra I class is scheduled for two periods or 90 minute per day and is designed for students who have below average math scores according to the Grade 8 Explore test. The additional 45 minutes per day is intended to extend instructional time to allow for increased reinforcement of skills and differentiation. The participating high schools schedule their 2P classes differently. Diverse HS scheduled the two algebra periods consecutively; students are released for a five-minute passing period between periods and Suburban HS separated the two periods, scheduling the second period of Algebra 1 later in the day.

The average class size for all participating classes was 25.8 students. As can be seen in Table 5, mean class size differed by school and by class type with smaller classes evident at Suburban HS and in 2P classes. Diverse HS served on average 29% more students than Suburban HS in each Algebra 1 class the equivalent of more than eight additional bodies. The difference in class size was even more evident among 2P between the two high schools with an average of more than 10 additional students in Diverse HS 2P classes compared to Suburban HS, a 37% difference.

Table 5

Average class sizes in participating classrooms by school and class type.

	Diverse HS		Suburban HS		Both HS
Class Type	n	Class Size	N	Class Size	Average
1P	3	32.6	3	25.5	29.0
2P	3	27.6	3	17.3	22.5
All Types	6	30.1	6	21.5	25.8

Note. Single-period class type= 1P; Double-period class type=2P.

Actual attendance rates on observation days differed between the two schools with higher rates of attendance evident at Suburban HS (91.5%) compared to Diverse HS (84.5%). Attendance findings are consistent with data gathered from school report cards, shown in Table 6, indicating higher rates of truancy at Diverse HS than Suburban HS.

Table 6

Mean actual observation day attendance rate and expected based on class enrollment at each high school by class type.

	[c]Single-Period (1P)		[d]Double-Period (2P)	
School/Class Type	Actual	Expected	Actual	Expected
[a]Diverse High School	27.0	32.6	24.4	28.3
[b]Suburban High School	23.8	26.0	18.7	20.6

Note. [a]n=18; [b]n=18; [c]n=12; [d]n=24.

Teacher Participants. All Grade 9 algebra teachers from two district high schools (n=18) were invited to participate in the observation study through an email describing the study along with a copy of a *Consent to Participate Form* (See Appendices B and C). Of the 17 consenting teachers, 15 were selected from two high schools to achieve the target of 12 classrooms (*n*=7, Diverse HS, *n*=8, Suburban HS) with three 1P and three 2P classrooms at each school. Based on almost total participation it may be said that the behaviors observed within this sample is likely to be representative of typical local practices.

The email and the consent form communicated that participation is voluntary and a decision to decline would not result in negative consequences. Teachers were assured of strict confidentiality. Within one week of the email invitation, the researcher followed up with an additional email and a telephone call to potential teacher participants to answer questions and confirm study interest. Teachers who did not wish to participate were instructed to not return the consent form. Classrooms were excluded if a teacher or co-teacher declined to participate in the study.

Teachers were also informed about a possible benefit of participating in this study. Upon completion of the study, teachers will receive a results summary of their classroom observations indicating the degree to which included evidence-based practices were observed. Such information has the potential to increase teacher self-awareness about their current instructional practices and professional development needs.

Demographic information, including education and teaching experience was collected on the selected teachers and summarized in Table 7. All but one of the participating teachers was female. Teachers participants from Diverse High School were all Caucasian females between the ages of 25 and 45. Suburban HS teachers were also Caucasian and predominantly female (one male), and between the ages of 24 and 40.

Student. Given the variability in attendance day to day and the fact that the primary participants are the teachers, limited data was collected on students. For a single observation session for each class type, the number of students in attendance was tabulated to generate average class sizes.

Table 7

Teacher participant's educational attainment and years of experience

High School	Certification	Educational Attainment	Mean years of Experience
Diverse HS	General	2 B.A.	
		4 M.A.	6
	Special	1 M.A.	30
Suburban HS	General	4 B.A.	
		2 M.A.	5
	Special	2 M.A.	10

Note. B.A= Bachelor Degree; M.A.= Master's Degree. Mean years of experience for all teacher participants=8.

Researcher and Observers. The researcher and primary observer in the study was an employee of the participating district at one of the participating high schools. The researcher was not and had never been in an administrative or evaluative role in the district. A second observer was a School Psychologist recruited from one of the participating high schools. In addition, a School Psychology practicum student recruited from the researcher's university acted as a third and final observer. None of the observers employed in the study had evaluative capacity over participating teachers.

Development of the Observation of Evidence-Based Instruction Instrument

The observational instrument used in this study, the *Observation of Evidence-Based Instruction (OEBI)* was developed based on Cornett's (2010)

Classroom Observation Scale (COS) that was used in his study of secondary-level general education evidence-based instructional practices. See Appendix D for a copy of the *COS*. Cornett created the *COS* beginning with a comprehensive literature review for evidence-based instructional practices appropriate for secondary classrooms and three editions of the *Handbook of Research on Teaching* (Gage, 1965; Richardson, 2001; Wittrock, 1986). From this review, 142 instructional and management activities were identified and a brief description was generated of each activity. These separate activities then were organized into 30 categories. This preliminary organization then was reviewed by a nine-member expert panel, each of whom had extensive background in high school educational research and practice. Based upon a literature review and the expert advice, Cornett's (2010) final version of *COS* was organized into the following four principal activities (a) Student On-Task, (b) Learning Arrangement, (c) Transition Time, and (d) Instructional Activity.

For this study, the *COS* was modified so that the primary focus is on the two activity types, Instructional Activity, and Transition Time. The Student On-Task and Learning Arrangement activities were omitted because the focus of this study is on teacher behavior. Transition Time was included because it contributes to understanding potential time not spent engaged in instructional activities. However, Transition Time was recorded only when transitions last 30 seconds (an entire observation interval). Instructional Activity also included evidence-based instructional practices with the highest effect sizes in the meta-analyses by Marzano et al. (2001) and Hattie (2009) and teaching practices specific to high quality algebra instruction. To be consistent as possible with Marzano et al. and Hattie's highest effect size practices and labeling, some of the *COS* instructional categories were renamed, some subcategories were dropped, and some new instructional categories and subcategories were added. The final version of the OEBI used in this study is provided in Appendix E.

Development of Instructional Activity. The final version of Instructional Activity in the OEBI consists of three EB categories: (a) EB Strategy Instruction, (b) Feedback, and (c) Formative Assessment. Practices not considered EB were also included in the OEBI to account for time spent not engaged in EB practices and included time spent in (a) Low or No Effect Size Practices, (b) Not Engaged in Instruction, and (c) Task Management Activities. A summary of all changes and modifications to EB practices from the COS to the OEBI is presented in Table 8. The origin of each subcategory and the rationale for altered and added subcategories is included. The operational definitions for the subcategory variables to be observed using the observation instrument are found in Appendix F. Many of the definitions are derived directly from the COS. Any changes to original definitions are also explicitly marked.

Table 8

OEBI Evidence-Based (EB) Instructional Practices of Interest, Changes from Cornett's (2010) COS, and Rationale for Inclusion- Rank and Effect Size.

Instructional Practices	Changes from COS	Marzano et al. (2001) Label & Rank (ES)	Hattie (2009) Label & [a]Rank (ES)
EB Strategy Instruction			
Teacher Think-Aloud	Renamed from Explicit Modeling.	1st Similarities & Differences (1.61)	12th DI Component (.57)
Explicit Review	Consolidated from three types of review into one.	2nd Summarizing (1.00)	12th DI Component (.57)
Guided Practice	New	4th General Practice (.77)	5th Spaced vs. Massed Practice (.71)
Independent Practice	New	4th General Practice (.77)	5th Spaced vs. Massed Practice (.71)
Student Think-Aloud	New	--	6th Self-verbalization (.69)
Self-evaluation	New	--	6th Metacognitive Strategies (.69)
Problem Solving Teaching	New	8th Generating & Testing Hypotheses (.61)	8th (.61)
Effort/ Persistence Reinforcement	New	3rd Reinforcing Effort & Providing Recognition (.80)	--
Visual or Graphic Device	Renamed from Graphic Devices and Organizers.	5th Nonlinguistic Representations (.75)	15th Concept Mapping (.57)

Instructional Practices (cont.)	Changes from COS	Marzano et al. (2001) Label & Rank (ES)	Hattie (2009) Label & [a]Rank (ES)
Worked Examples	New	--	14th (.57)
Mastery Goals	New	--	13th (.56)
Elaborated Feedback	Unchanged	7th Setting Objectives & Providing Feedback (.61)	4th Feedback (.73)
Formative Assessment			
Formal Progress Monitoring	Unchanged	--	1st Providing Formative Evaluation
Checks for Understanding	Renamed from Monitoring and Questioning	9th Questions, Cues, & Advanced Organizers (.59)	21st Questioning (.44)

Note. [a] Hattie rankings include only the effects in the domain of teaching practices.

-- Indicates an average effect size was not generated.

Evidence-Based Strategy Instruction. The Evidence-Based (EB) Strategy Instruction category encompassed 11 observable EB practices within three instructional subcategories including Direct Instruction (DI), Metacognitive Strategies, and Other Strategies. Several *COS* original categories were recast and altered as subcategories. For example, the original category of Modeling in the COS included two subcategories, Explicit Modeling and Implicit Modeling. In the OEBI, EB Explicit Modeling is named Teacher Think-Aloud (similar and incorrect examples) and Implicit Modeling is retained as a Low or No Effect Size Practice. The original COS category of Review was modified and renamed Explicit Review. The original three subtypes of Review, Review of

Facts/Concepts/ Procedure, Manipulate/Generalize, and Skill/Strategy were consolidated under the new Explicit Review subcategory.

Several EB practices are new to the OEBI, for example the Metacognitive Strategy subcategories. Student Think-Aloud, Self-evaluation, Problem Solving Strategy or Heuristic, and Reinforcement of Effort/Persistence were added to the OEBI based on high effect sizes reported in the Hattie (2009) syntheses of meta-analyses. Operational definitions of each original, adapted, and new instructional practice subcategories are found in Appendix A.

Feedback. The COS Feedback category containing Elaborated Feedback and Simple Feedback was retained in the OEBI.

Formative Assessment. The category for Formative Assessment was altered from the original Formal Assessment of Learning on the COS to maintain the focus on the highly effective evidence-based assessment practices. The original subcategory, Formative Progress Monitoring, was not changed. The OEBI Informal Checks for Understanding subcategories were renamed COS subcategories of Questioning for verbal, written, or action response. The original COS definitions were used with the exception that non-examples were added.

Non- Evidence-Based Instruction. Non-EB versions of several of the subcategories are included on the OEBI (Appendix E). Non-EB instructional practices fell into three subcategories including Low and No Effect Size Practices, Not Engaged Practices and Task Management Activities made up of nine discrete variables as shown in Table 9. For example, Implicit Modeling is marked when the teacher demonstrated a skill or strategy without verbal mediation and non-EB Lecture is scored when teachers verbally explain concepts without demonstration. Task Management Activities, like Giving Directions or Not-Engaged Practices like Physical Monitoring were included based on the possibility that they may account for significant portions of class time.

Table 9

Low Effect Size Practices, Not Engaged Practices and Task Management Activities

Low Effect Size	Not Engaged Practices	Task Management Activities
Simple Feedback	Not Engaged in Instruction	Giving Directions
Lecture	Transition Time	Managing Classroom Behavior
Problem-Based Learning	Physical Monitoring	
Implicit Modeling		

Post-Observation Teacher Report

After each full 45-90 minute observation, teachers completed a brief post-observation reflection sheet (See Appendix F), including information regarding the unit material (e.g., new, review, end of unit). In addition, teachers self-evaluated on a three-point scale (e.g. Atypical; Somewhat Typical; Very Typical) how representative the observed instruction is of a typical day’s instruction with the option to add a short explanation of the rating.

Scaling

During instructional time, the OEBI was completed via a partial-interval recording (PIR) method where the single occurrence of each instructional activity of interest was recorded during each interval (Cooper et al., 2007). The OEBI is divided into 30-second segments that become the time sampling intervals within which instructional activities were recorded if they were observed. Observed instructional activities were recorded once and only once in each sampling interval. For example, if Elaborated Feedback was observed four times within a 30-second interval, only one tally was recorded. All OEBI subcategories were recorded simultaneously. For example, if Explicit Modeling, Elaborated Feedback, and use of a Visual/Graphic Device all occur within a single 30-second interval, all three instructional practices were marked on the observation form in that interval. If none of the identified subcategories of evidence-based practice occur throughout an entire 30-second interval, the observer will record a tally in one of the Other Non-instructional Activities, or the Transition Time or Not Engaged in Instruction options.

The categories of Transition Time, Lecture, Physical Monitoring and Not Engaged in Instruction were Whole-Interval Recorded (WIR), meaning these activities are only recorded when they last the entire 30-second interval. WIR is the most conservative type of interval recording, and may underestimate the occurrence of a behavior. By using WIR for Transition Time, Lecture, Physical Monitoring, and Not Engaged in Instruction, the risk of overestimating non-instructional time is significantly reduced. For every observation interval a behavior were recorded, therefore accounting all activities that occur for the duration of every observed class period.

Observer Training

First, three observers were read and discussed the operational definition for each category and subcategory of instructional activity. Second, both observers practiced data collection using the observation form with a videotaped high school algebra lesson. Third, participating Algebra classrooms were co-observed until observers are in 90 percent agreement in each of the large categories. At this point the remaining classrooms were scheduled for data collection. Observers were assigned classrooms to observe within a four-week data collection window. The secondary observers each observed four different classrooms for the duration of the single or double period depending on class type. The primary researcher observed the remaining 28 classrooms.

Inter-observer Reliability

To determine inter-observer agreement, two data collectors were independently observe and score the time sample intervals for two full Algebra 1 class periods. Inter-observer percent reliability agreement was calculated using the following formula:

$$\frac{\text{Number of Agreements}}{\text{Number of Agreements}+\text{Disagreements}} \times 100 = \text{Percent Reliability}$$

(Martella, Nelson, Morgan, & Marchand-Martella, 2013). Once 90% inter-observer agreement is achieved the observers will scheduled formal data collection. One time per week of data collection, a classroom will be co-observed for 30 minutes to verify continued inter-observer agreement.

Procedures

All observations will be conducted within a 3-4 week period of time to increase the likelihood that similar material is covered across classrooms and similar external factors may be in play, to reduce contextual variability (e.g., impact of the academic calendar) holidays or other district events). Each of the 12 algebra classrooms will be observed twice for the duration of the class period (e.g. 45 or 90 minutes) by the researcher and/or observers.

Data Collection. A total of 36 45-minute periods of Algebra 1 were observed in 12 classrooms. Six single-period (1P) classrooms and six double-period (2P) classrooms were each observed twice. Five of the classrooms were co-taught with either a general educator and a special educator (*n*=4) or two general educators (*n*=1). Differences between the two participating schools in the allocation of staff resources were evident. For example, none of the 1P classes taught at Diverse High (0 out of 7) were co-taught, while more than half the 1P Algebra 1 classes at Suburban High (4 out of 7) were co-taught. The staffing of 2P Algebra 1 classes was more equivalent between high schools, but still favoring Suburban HS. Suburban HS staffed 75% of 2P Algebra 1 with a co-teacher (3 out of 4), while 57% of Diverse HS 2P classes were co-taught (4 out of 7). Participating classrooms were matched to maximize similarities, with the unavoidable exception that no co-taught 1P classes from Diverse HS were included.

The observer arrived prior to the start of class and sat in the back of the classroom where they did not interfere with instruction but could see every student. Observer(s) had an electronic copy of the observation sheet on an I-Pad and a back-up paper copy, in case of technical difficulties. A 30-second repeating countdown clock was positioned near the observer(s). A head count was taken at the 15-minute mark to account for tardy students. The 30th interval is highlighted as a reminder to collect this data. Data collection was conducted in real-time beginning when the bell rings and ending when the bell rings. The observer did not record any instructional activity that occurred after the formal beginning or end of the class period. For the duration of each interval the observer recorded all instructional activities as they occurred. The repeating countdown clock automatically reset to 30 seconds after each interval at which time the instructional activity was recorded in the next row down. This process repeated until the class ends. In 2P classes two observation sheets were used and coded as 2P-1 or 2P-2 to signify the first period and second period of the class.

In co-taught classrooms, the observer(s) will be followed the behaviors of both teachers simultaneously. All instructional activity was recorded regardless of which teacher provided the instruction. For example, if one teacher was engaged in Explicit Review, while the other teacher was providing Elaborated Feedback to a student, both activities were recorded. However, in cases where one teacher was engaged in an instructional activity and the other is Not Engaged in Instruction (e.g. checking emails), only the instructional activity is marked. For this category to be marked in a co-taught classroom, both instructors had to be Not Engaged in Instruction.

Data Analysis

Observation data was analyzed in three ways. First, the percentage of total instructional time spent engaged in highly effective instructional practices was determined. In order to determine the percentage of time spent engaged in high effect-size EB practices compared to non-essential (i.e., Low or No Effect Size Practices, Not Engaged in Instruction, and Task Management Activities) practices, the number of intervals in which each subcategory was observed was divided by the total number of observation intervals.

$$\frac{\text{\#of intervals observed EB practice}}{\text{Total \# of observation intervals}} = \text{Percentage time engaged in EB practices}$$

Second, the percentage of total instructional time spent engaged in each specific instructional practice was calculated. In order to determine the total percentage of instructional time spent engaged in highly effective, evidence-based (EB) instructional practices, the number of intervals marked for any identified EB subcategory will be divided by the total number of intervals in the observation.

$$\frac{\text{\# of intervals subcategory is observed}}{\text{Total \# of observation intervals}} = \text{Percentage time engaged in subcategory}$$

Because discrete instructional practices were recorded simultaneously, percentages were not expected to add up to 100%, therefor the percentage of intervals in which only EB practices were observed, only non-EB practices were observed, and instances where both an EB and a non-EB practice were observed simultaneously were calculated.

Third, to determine the significance of differences in observed instructional practices as a function of student achievement level, 16 discrete EB instructional practices observed in the study were grouped logically by similarities to form three dependent variables, (a) Evidence-Based Strategy Instruction, and (b) Feedback, and (c) Formative Assessment, as shown in Table 10. For example, EB Strategy Instruction was composed of the instructional practices under Direct Instruction Total, Metacognitive Strategies, and Other Strategies and Formative Assessment included Formal Progress Monitoring and Checks for Understanding. The exception, (b) Feedback consisted of only Elaborated Feedback. A one-way MANOVA was conducted to determine if differences in the three dependent variables were evident by 1P or 2P classes.

Table 10

Component instructional practices for the three EB dependent variables.

Dependent Variable	Evidence-Based Strategy Instruction	Formative Assessment	Feedback
Included Instructional Practices	Explicit Modeling Explicit Review Guided Practice Independent Practice Student Think-Aloud Self—Evaluation Problem-Solving Teaching Reinforcement of Effort/Persistence Visual or Graphic Device Mastery Goal Worked Example	Formal Progress Monitoring Verbal Checks for Understanding Written Checks for Understanding Actions Checks for Understanding	Elaborated Feedback

CHAPTER 4- RESULTS

The purpose of this study was to observe systematically the instructional methods that Grade 9 high school algebra teachers use and the degree to which these methods included evidence-based practices. The goal was to understand typical mathematics algebra instructional practices with the potential to lead to targeted staff development, should there be a deficit in these evidence-based practices.

It was hypothesized that Grade 9 Algebra 1 instruction does not regularly employ the most powerful evidence-based practices as identified consistently in the professional literature, such as Evidence-Based Strategy Instruction, Formative Assessment, or Elaborated Feedback. Secondarily, it was hypothesized that there would be differences in frequency or types of instructional practices (IPs) utilized in classes serving average achieving students versus classes serving below average achieving students.

Establishing the Validity of the Independent Variables

Fundamental to interpreting the study's results is an assumption that the observed algebra teachers' instruction was representative of their everyday instructional practices. The possibility that the observed instructional practices were unrepresentative may not be apparent to an external observer and may threaten the accuracy of drawn conclusions (Shadish et al., 2002). To ensure that the observed instruction reflected typical instruction, the opinions of the participating teachers were deemed critical. Therefore, participating teachers filled out a post-observation questionnaire to determine the degree to which they believed their observed instruction was representative of their typical practices. A copy of this teacher self-rating is included in the appendix. The typicality of observed instruction was rated from 1 (*very atypical*) to 4 (*very typical*) scale. As shown in Figure 2, nearly all teachers (94%) reported that their observed instruction was *very typical* or *somewhat typical* of what and how they taught every day. Only two teachers (6%) reported instruction was *somewhat atypical* and no teacher claimed instruction was *very atypical*. Based on these results, it is likely that the data collected represents an accurate sampling of typical instruction in the observed classrooms, improving chances that the observed instruction may be interpreted as a generalizable pattern over other times, if not other persons or settings.

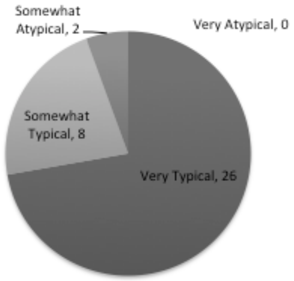

Figure 2. Post-observation teacher self-ratings verify observed instruction typical of regular practice.

Instructional Practices Observed

Descriptive statistics from observational data collected with the *OEBI* from 36 Algebra 1 class periods are shown in Table 11. Observed instructional practice (IP) means, medians, ranges, and standard deviations are shown along with Hattie's (2009) or Marzano et al. (2001) calculated effect size, if available. These data also are displayed graphically in Figure 3.

Table 11

Descriptive statistics for the percentage of time spent in instructional activities, including previously reported effect sizes.

Instructional Activity	Mean	Median	Range	SD	ES
Direct Instruction Total	63.6	71.0	76.0	21.9	[a].59
Explicit Modeling	9.4	6.5	53.0	11.4	[a].59
Explicit Review	6.3	6.0	19.0	5.9	[a].59
Independent Practice	25.2	22.5	89.0	23.3	[b].77
Guided Practice	22.7	24.0	56.0	17.3	[b].57
Metacognitive Strategies Total	4.4	3.0	19.0	5.4	[a].67
Student Think Aloud	2.9	1.0	14.0	4.2	[a].64
Self-Evaluation	0.1	0.0	11.0	1.9	[a].62
Problem Solving Teaching	0.1	0.0	6.0	1.3	[a].61
Reinforcement Effort/Persist.	0.4	0.0	3.0	0.8	[b].80
Other Strategy Instruction Total	4.6	4.0	17.0	4.9	[a].60
Visual or Graphic Device	3.3	0.0	16.0	4.5	[b].75
Mastery Goal	0.7	0.0	9.0	1.8	[a].58
Worked Example	0.7	0.0	6.0	1.3	[b].59

Instructional Activity	Mean	Median	Range	SD	ES
Elaborated Feedback	17.7	15.0	42.0	14.1	[a].73
Formative Assessment Total	18.8	15.5	40.0	12.3	[a].59
Verbal Check for Understand.	15.2	12.0	42.0	12.1	[b].59
Written Check for Understand.	2.9	0.0	32.0	6.9	[b].59
Action Check for Understand.	0.6	0.0	22.0	3.6	[b].59
Formal Progress Monitoring	0.1	0.0	1.0	0.2	[a].90
Low Effect Size Practices	11.6	--	--	--	NA
Simple Feedback	6.8	7.0	18.0	5.4	[b].22
Lecture	4.0	3.0	13.0	3.8	NA
Problem-Based Learning	0.0	0.0	2.0	0.3	[a].15
Implicit Modeling	0.8	0.0	12.0	2.1	NA
Not Engaged Total	15.8	--	--	--	NA
Not Engaged in Instruction	8.3	8.0	21.0	6.50	NA
Transition Time	5.3	4.0	14.0	4.3	NA
Physical Monitoring	2.2	0.0	11.0	3.5	NA
Task Management Activities	11.1	--	--	--	NA
Giving Directions	6.4	6.0	17.0	4.2	NA
Managing Classroom Behavior	4.7	3.0	23.0	5.3	NA

Note. Instructional practices in *Italics*=Whole interval recorded. All other instructional variables could co-occur with other variables.

Mean, median and range indicate percentage of observed intervals

[a] ES taken from Hattie (2009)

[b] ES taken from Marzano et al.(2001)

– indicates value not calculated

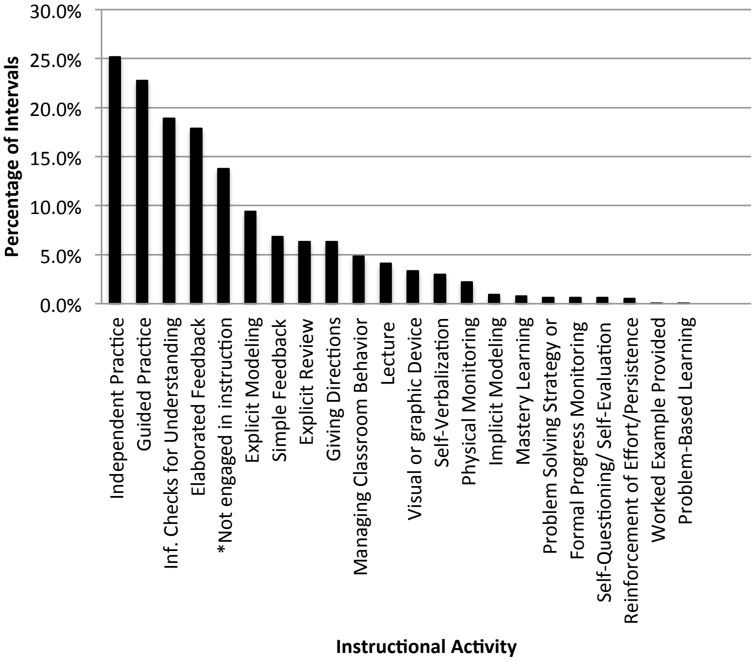

*Figure 3***.** Graphic representation of observed instructional variables arranged from most to least frequently observed.

In contrast to Cornett's (2010) study that showed the most common practice observed during class time was Not Engaged in Instruction, in this study, the Evidence-Based Strategy Instruction group Direct Instruction Total was the most common practice, observed in more than two-thirds of the intervals (median = 71.0). When Direct Instruction Total was combined with all other categories of instruction, including Metacognitive and Other Strategy Instruction, Feedback, Formative Assessment Total, Low Effect Size Practices, and Task Management Activities, results indicated that teachers were engaged in some type of instructional practice during nearly *all* of the observed intervals (84.2%) of class time. Very little time was spent in non-instructional activity (15.8%), with most of that time spent Not Engaged in Instruction (8.3%) or in Transition (5.3%). Physical Monitoring made up another 2.2% of intervals.

More than three-quarters of Direct Instruction Total (mean = 63.6%) consisted of some form of student practice. Practice time was divided almost equally between Independent Practice (mean = 25.2%) and Guided Practice (mean = 22.8%) combining to account not only for the majority of Direct Instruction Total, but for almost half of all observed intervals.

Beyond the practice components of Direct Instruction Total, the seven types of Strategy Instruction were observed relatively infrequently or not at all making up 7% of instructional intervals. Specifically, Metacognitive Strategies Total (median=3.0%) and Other Strategy Instruction (median=4.0%) were rarely observed.

High variability of instructional practice frequency between classrooms also stands out as a particularly striking finding. For almost every instructional variable, the standard deviation approached or exceeded the mean. As can be seen in Table 3, high variability was observed in every variable but Direct Instruction Total. Typical of this pattern, Metacognitive Strategies Total was observed in an average of 4.4% of intervals, but the standard deviation was 5.4%. Such variability was indicative of a pattern whereby an IP was observed 0% in 12 classrooms, but between 4% and 19% of intervals in half of observed classrooms, as with Metacognitive Strategies Total. Consistently large individual teacher differences in frequency of IP use suggest that such variability was the norm.

Frequency of EB instruction

Of total time, 63% was spent only engaged in practices considered EB, meaning no non-essential instruction (e.g. Task Management Activity or Low or No Effect Size Practice) occurred in the same interval. Figure 4 depicts the percentage of time spent only in EB practice compared to non-EB practices. Both an EB practice and a non-EB practice were recorded simultaneously in only 10% of intervals. Results suggest that EB instruction occurred in 73% of intervals. The overall results presented a positive picture of current Algebra 1 instructional practice with almost three fourths of class time falling into at least one identified highly effective category.

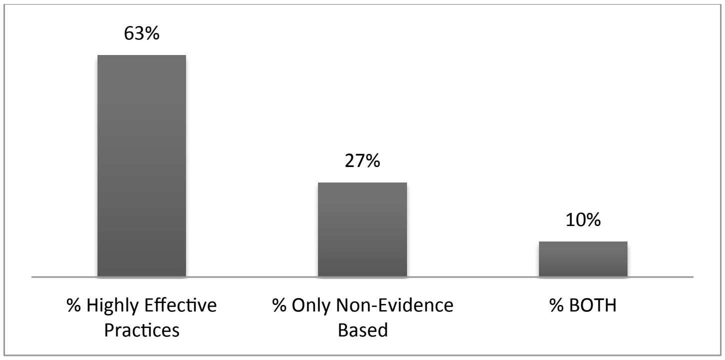

Figure 4. Percentage of time spent in only EB practice, non-EB practice, and both

To simplify understanding of the frequency of IPs that are evidence-based or “highly effective”, Table 11 variables were ordered from largest to smallest average effect size in Table 12. Several IPs that did not have an identified effect size (e.g. Implicit Modeling or Lecture), as well as Task Management Activities, were not included in Table 4.

Table 12

Instructional practice listed in order from highest to lowest effect size with associated mean observed frequency.

Instructional Practice	Mean	ES
Formal Progress Monitoring (FA)	0.1	[a].90
Reinforcement of Effort (SI)	0.4	[b].80
Independent Practice (SI)	25.2	[b].77
Guided Practice (SI)	22.7	[b].77
Visual/Graphic Devices (SI)	3.3	[b].75
Elaborated Feedback (FB)	17.7	[a].73
Student Think Aloud (SI)	2.9	[a].64
Self-Evaluation (SI)	0.1	[a].69
Problem Solving Teaching (SI)	0.1	[a].61
Explicit Modeling (SI)	9.4	[a].59
Explicit Review (SI)	6.3	[a].59
Verbal Check for Understanding (FA)	15.2	[b].59
Written Check for Understanding (FA)	2.9	[b].59
Action Check for Understanding (FA)	0.6	[b].59
Simple Feedback (LE)	6.8	[a].22

Problem-Based Learning (LE)	0.1	[a].15

Note. ES= Effect Size; SI= Strategy Instruction; FB=Feedback; FA=Formative Assessment; LE= Low Effect Size Practices

[a] ES from Hattie (2009)

[b] ES from Marzano et al. (2001)

Although evidence-based IPs were frequently observed, only a few of the top-ranked highly effective IPs were observed to occur in more than 10% of the observed intervals; many highly effective IPs were observed infrequently (e.g., Reinforcement of Effort) or not at all. For example, Formal Progress Monitoring, the top-ranked highly effective IP was observed in only one classroom, making up only .1% of all observation intervals.

As discussed under Question 1, the Evidence-Based Strategy Instruction group Direct Instruction Total (DI) was observed frequently in Algebra 1 classrooms making up more than two-thirds of all instruction with most of that time spent in Guided and Independent Practice. As seen in Table 5, Independent and Guided Practice are highly effective strategies. However, only about a quarter of observed DI consisted of key components teacher-led Explicit Modeling (9.4%) and Explicit Review (6.3%) of skills and strategies. Teachers typically spent less than a sixth of class time explicitly modeling skills and strategies, despite the fact that 87% of class time was typically devoted to new material and review of previously learned material according to the Post-Observation Teacher Questionnaire.

The lack of observed Evidence-Based Strategy Instruction other than a form of practice was one of the most striking findings. Although seven strategies make up Metacognitive Strategies and Other Strategy Instruction, these seven strategies combined accounted for less than 9% of instructional time. In addition, just two strategies, Student Think-Aloud (2.9%) and the use of Visual and Graphic Devices (3.3%) made up for most of that small amount of Strategy Instruction. Visual and Graphic Devices, the most commonly observed strategy was not used at all in more than half of all classrooms (*n*=20) and the use of concrete representations or manipulatives was never observed. Other than the practice components of DI, only two other highly effective IPs were observed close to 20% of the time, including Verbal Checks for Understanding (15.2%) and Elaborated Feedback (17.7%).

Verbal Checks for Understanding accounted for 81% of all Checks for Understanding. Verbal Checks for Understanding were observed most frequently when directed at individual students (e.g. “Juan, is the correlation positive or negative?”) during whole group instruction. Teachers rarely pulled for a group or choral response. Other Checks for Understanding types (e.g., Written and Action Checks for Understanding) were rarely observed. The data suggests that Formative Assessment practices consisted primarily of teacher questioning and verbal response.

Elaborated Feedback occurred on average in 17.7% of observed intervals most frequently during time spent in Guided Practice. Elaborated Feedback was observed more than twice as often as Simple Feedback (6.8%), a Low Effect Size Practice. Teachers were therefore more likely to provide extended explanations than to simply indicate if a solution was right or wrong. Although Elaborated Feedback was one of the more frequently observed IPs, as with the majority of IPs, the amount of Elaborated Feedback provided by teachers was highly variable (SD=14.1). Elaborated Feedback was observed in 0% to 41% of all intervals. A quarter of the observed classrooms (*n*=9) engaged in Elaborated Feedback frequently, in more than 31% of intervals. On the other hand, another quarter of observed classrooms engaged in Elaborated Feedback for fewer than 4% of the intervals.

Instructional Practice Differences by Class Type

As noted earlier in Chapter 3, in the participating high schools, students were placed in either a one-period (1P) algebra class or two-period (2P) algebra class based on their achievement levels. Students meeting the district criteria for *average* math skills were placed in a 1P algebra class. Students who earn *below average* mathematics scores were placed in a 2P algebra class, with the intent that an additional 45 minutes of intervention time would close the achievement gap.

A one-way MANOVA then was conducted to determine if differences in the three dependent variables were evident by 1P or 2P classes. To provide a clear picture of the differences in the three variables in 1P and 2P classrooms and by school, descriptive statistics are shown in Table 12.

Table 12

Summary of descriptive statistics for the combined Instructional Practice variables.

	EB Strategy Instruction	Formative Assessment	Feedback
Group (n)	M (SD)	M (SD)	M (SD)
IP	63.9	19.7	18.1
(n=12)	(16.3)	(9.9)	(14.0)
2P	66.2	16.5	24.3
(n=12)	(16.9)	(11.1)	(10.5)

Note. M=mean; SD=standard deviation.

The descriptive statistics suggests minimal differences between 1P and 2P (e.g., mean EB Strategy Instruction 1P=63.9, 2P=66.2) classes suggesting that instructional practice frequencies may not vary reliably based on the achievement level of students or enrollment in one or two periods of Algebra 1.

MANOVA results are found in Table 13 corroborating the surface-level analysis that were skeptical about 1P vs. 2P differences, as 1P vs. 2P interaction effects were not found. Although differences in instructional practices were initially expected between 1P and 2P classes, multivariate main effects were *not* detected, F (3,18) = 1.2, p = .343.

Table 13

Multivariate analysis of variance (MANOVA) results for evidence-based instructional practices according to group membership.

Group	Wilks' Lambda	F	Hypoth df	Error df	Sig.
1P vs. 2P	.835	1.186	3	18	.343

Note. ***p*=<.01. 1P= single-period class; 2P= double-period class.

Table 14

Univariate tests of between-subjects effects 1P vs. 2P.

Group	Dependent Variable	F	Sig.
1P vs. 2P	EB Strategy Instruction	.138	.714
	Feedback	1.417	.248
	Formative Assessment	.823	.375

Note. **p*<.05; ***p*<.01. EB=Evidence-Based.

Although the 2P class is in place to better address the needs of math students entering high school without the pre-requisite skills to be successful in Algebra 1, it appears that other than extended instructional time, instructional practices are not different in type or frequency percentage.

CHAPTER 5- DISCUSSION

The purpose of this study was to observe systematically the instructional methods that Grade 9 high school algebra teachers use and the degree to which these methods include evidence-based practices. The international and national comparison data established the ongoing and widespread problem of US student math under-achievement, especially at the secondary level when students make the transition from arithmetic to algebra. The pervasive lack of algebra readiness for U.S. high school students paired with the established importance of Algebra 1 creates a dilemma for American secondary educators. It is incumbent on Grade 9 Algebra 1 teachers to provide curriculum and instruction that prepares students for future even more challenging course work. The curriculum and instruction at the high school level must therefore be highly effective. According to 40 years of research, three categories of instructional practices are consistently shown to enhance student achievement in math and across subject areas. These often overlapping methods, each of which can be said to improve teacher clarity, include (a) EB Strategy Instruction, (b) feedback, and (c) formative assessment. It is the hypothesis of this study that teachers do not frequently use EB instructional practices and that student math achievement could be significantly upgraded if teachers recognize the potential power of evidence-based (EB) practices to enhance their instruction.

Five major conclusions can be drawn from this study with associated implications for teachers, instructional practices, future service delivery models, and future research.

1. High school Algebra I teachers were (a) engaged in instruction and (b) spent little time on non-math or non-teaching activities (e.g. checking email, transitions, etc.) more than might be predicted based on prior research.

2. Instructional practices among teachers were highly variable.

3. Most instructional time was spent in activities categorized as evidence-based, but a preponderance of that teaching time was spent in two forms of practice (Guided and Independent) with little time devoted to other EB Strategy Instruction.

4. Informal Formative Assessment often lacked variety and depth.

5. No reliable differences in instructional practices were noted in 1P and 2P classes, even though they were intended to reflect differences in students' instructional needs.

Teachers teaching most of the time. Contrary to expectations based on prior research (e.g.,Burns, 1984; Gump, 1967), study results indicated that teachers *were* engaged in some form of instruction during nearly all of the observed intervals (86.2%) of algebra class time. This observed outcome contrasts to the reported results by Cornett (2010), who found teachers were Not Engaged in Instruction 23.2% of the time, almost triple the 8.3% found in this study.

Prior studies also indicated that significant amounts of instructional time are often lost to inefficient transitions between activities with some studies reporting an average of 15% per class period (e.g.,Burns, 1984; Gump, 1967). In this study, teachers spent very little (4%) allocated instructional time in Transitions.

The finding that teachers are primarily engaged in instruction rather than disengaged or losing time to transitions may possibly allay fears that a large amount of instructional time is wasted due to teacher disengagement or transitions. Such a finding allows the discussion to focus on instructional content and quality rather than a lack of instruction.

Instruction varies from one class to another. For almost every instructional variable other than Direct Instruction Total (Mean=63.9%, SD=21.9%), the frequency of use standard deviation approached or exceeded the mean, indicating that instructional practices varied considerably from one class to another. For example, Student Think-Aloud (Mean=2.8%, SD=4.2), was never observed (0%) in half of all classrooms, but was noted between 2% and 12% of intervals in the remaining half of observed classrooms. Similarly, Visual and Graphic Devices (Mean=3.3%; SD=4.5), were never used in more than half (20) of the classrooms, used between 4% and 7% in 10 classrooms and as much as 16% of intervals in one classroom.

At the classroom or teacher level, these differences were especially apparent. It is important to note, again, that frequencies will not add up to 100% because all instructional practices occurring in an observation interval were recorded simultaneously, with the exception of Not Engaged in Instruction, Transition Time, Lecture, and Physical Monitoring, which were recorded by whole-interval (see Methods).

The variability of instructional practice use frequency use can be illustrated by comparing individual teacher practices in two double-period (2P) classes. Teacher A and Teacher B engaged in similar high percentages of Independent Practice, but differences in frequency of other instructional practice use were stark. Teacher A spent 59% of total time in Independent Practice, provided no Feedback (0%), and very little time was spent in Checks for Understanding (8%). In contrast, Teacher B, teaching the same content, devoted similar amounts of time to Independent Practice (56%), but actively engaged in Checking for Understanding (41%) and providing Elaborated Feedback (33%) during the observation. Teacher B also used Explicit Modeling and Review practices another 39% of instructional intervals, nearly seven times more often compared to the first teacher (6%). Although both teachers engaged in EB practice for a large percentage of class time, differences in instruction during Independent Practice suggest differences in the quality of practice.

High variability of instructional practice use from one class to another is consistent with NMAP's (2006) conclusion that the existence of "substantial differences in the mathematics achievement of students are attributable to differences in teachers" (p. 51), that relates to individual teacher instructional decisions and use of evidence-based practices. Although such variability may be, in part, attributable to purposeful, flexible instructional decision-making based on ongoing formative assessment, perhaps it is also indicative of a lack of common algebra instructional goals or focus on *how* we teach algebra in the district or the school. The notion of common planning and assessment that has the power to improve teacher effectiveness on a large scale may be at odds with beliefs about the need to preserve teacher choice and independence.

> Americans hold the notion that good teaching comes through artful and spontaneous interactions with students during lessons. This kind of on-the-fly decision-making is made possible by the innate intuitions of "natural" teachers. Such views minimize the importance of planning increasingly effective lessons" (Stigler & Hiebert, 1997, p. 21).

Studies suggest that teacher's beliefs about loss of autonomy may be replaced as "a shared sense of intellectual purpose and a sense of collective responsibility for student learning" develops and achievement gaps narrow (Darling-Hammond et al., 2009, p. 11). As teachers work with one another to identify instruction that works, instructional practice variability may decrease in the pursuit of higher student achievement.

Most instructional practices were a variation of practice. Another consistent finding was that teachers spent a large percentage of instructional time engaging students in Independent Practice (25.2%) and Guided Practice (22.8%). Combined Independent and Guided Practice accounted for nearly half (48%)of instructional time, occurring more frequently than all other EB Strategy Instruction categories combined (27%). Practice time vastly overshadowed other powerful instructional practices.

Guided Practice and Independent Practice are two highly effective components of Direct Instruction (DI), a well-known teaching method that also involves Explicit Modeling and Explicit Review. DI accounted for 63.6% of total intervals and an even greater percentage of time spent engaged in instruction (85%).

According to the Marzano et al.'s (2001) analysis of the effects of practice, mastering a skill requires a significant amount of practice; students are shown to reach 80% mastery only after as many as 24 repetitions indicating that much practice is, in fact, needed for most learners. However, research also indicates that students when practicing are in a process of shaping their knowledge and deepening their conceptual understanding of a skill. When students have inadequate conceptual understanding, they may practice skills incorrectly or apply skills in a shallow manner (Mathematical Science Education Board, 1990). It is during this phase of learning that Feedback and Checks for Understanding are critical to success, but Elaborated Feedback (17.7%) and Checks for Understanding (18.8%) were in evidence for only a portion of practice time.

This raises concerns about too much or poorly timed practice US teachers are reported to spend too little time in the shaping phase of practice moving too quickly to a heavy independent practice schedule (Healy, 1990). The TIMMS video study showed that US algebra teachers typically engage in more independent problem solving practice than comparison countries, averaging 60% of instructional time (Givvin, Hiebert, Jacobs, Hollingsworth, & Gallimore, 2005). It is possible that not all of the Algebra 1 students in observed classrooms were ready to spend more than a quarter of class time practicing individually without Feedback or Checks for Understanding. Especially in 2P classes serving students known to lack pre-requisite Algebra 1 skills, it may be even more important that practice is consistent with instructional level.

Because practice made up three-quarters of DI Total, much less time was spent in proactive teacher-led Explicit Modeling (9.4% of total intervals) and Explicit Review (6.3% of total intervals) of skills and strategies. Furthermore, Explicit Modeling of non-examples that clarify for students when a strategy is used and not used (Archer & Hughes, 2011; Barbash, 2011) was almost never implemented (Mean=0.3%, SD=0.8). In fact, Marzano et al.'s most effective strategy (ES = 1.61) involves the Explicit Modeling of Similarities and Differences, a strategy that emphasizes the process of identifying both examples and non-examples of a concept or skill. The use of non-examples was rarely observed and when observed, only briefly (one or two intervals).

In addition to the paucity of Explicit Modeling and Review, the seven other types of Evidence-Based Strategy Instruction (i.e., Metacognitive Strategies and Other Strategies) were observed even less often, accounting for 7% of total intervals. For example, use of Visual and Graphic Devices was observed in only 3.3% of intervals. Out of 36 classrooms, more than half (56%) never used a Visual or Graphic Device and concrete representations were never observed. Other examples of rarely used EB strategies include Mastery Learning and Worked Examples, both highly effective instructional practices that are particularly relevant to mathematics. An emphasis on mastery goals was evident in only two classrooms and was never observed in 25 classrooms (69%). If learning to mastery was a goal of instruction, it was not clear among participating classrooms. Similarly, Worked Examples, (i.e., the provision of a successfully completed problem for use during guided or independent practice) were not observed in 95% of classrooms.

As shown earlier, metacognition is integral to effective mathematical problem solving (Garofalo & Lester, 1985; Schoenfeld, 1985, 1987, 1989, 1992) and yet Metacognitive Strategy Instruction including Student Think-Aloud, Self-Evaluation, Reinforcement of Effort/Persistence, and Problem Solving Teaching, was in sum observed in only 4.4% of intervals. On average less than 2 minutes per period of instructional time was devoted to these highly effective strategies. For example, Student Think-Aloud, where students were required to verbalize their thought process, was observed in only 2.9% of intervals. Half of all observed classrooms (18) had teachers who never required Student Think-Aloud. Student Think-Aloud was observed to be a routine practice for only one teacher who engaged in the practice for 14% of intervals both times observed.

Reinforcement of Effort/Persistence was rarely observed (.4%), on average less than 1% of intervals. This finding stands in contrast to the emphasis in the new Common Core State Standards for Mathematics (2010), where students are expected to "Make sense of problems and persevere in solving them." Over the course of 36 classroom observations, instances of explicit Reinforcement of Effort were observed in only 10 classrooms and in no classroom more than three times.

The sporadic use of explicit EB Strategy Instruction other than practice is not unique to the current study, and has several implications for practice. Prior studies also suggest that U.S. teachers tend to spend too little time using explicit instructional practices, such as Explicit Modeling or Visual and Graphic Devices, that if used have the potential to shorten learning time (Carnine et al., 1994; Schumaker et al., 2002). Despite research showing math teachers trained in heuristic methods of problem solving realize better student achievement outcomes (Hattie, 2009; Hattie et al., 1996; Marcucci, 1980) and although teachers report that they believe that teaching strategies is as important as teaching content, observation studies show that teachers rarely teach components associated with strategies in their class (Schumaker et al., 2002). The current study adds to these findings, pointing to the need to determine the reasons EB explicit strategy instruction remains more ideal than real. The dissemination of research-based practices to teachers for routine use continues to be problematic and deserves the focused attention of educators at all levels.

Formative assessment practices lacked variety and depth. Hattie (2009) insists that excellent teachers must at all times seek to answer three basic assessment questions that guide instruction:

1.“Where are they going?”

2.“How are they going?” and

3. “Where to next?”

Formal Progress Monitoring is a Formative Assessment practice that consistently shows very strong effects on achievement. Formal Progress Monitoring was only observed with one student in one classroom. Formal Progress Monitoring is enhanced by the self-graphing of results, scored as a form of the Metacognitive Strategy, Self-Evaluation. Students were observed to record/graph their own progress in only one classroom.

Instead of Formal Progress Monitoring, informal Checks for Understanding, such as questioning for verbal or written response, were the most frequent method (18.9%) to assess student knowledge. In fact, informal Checks for Understanding were the third most frequently observed teacher instructional practice after Guided and Independent Practice. By way of comparison, Cornett (2010) observed less than half the frequency of informal

Checks for Understanding. Theoretically, the higher the frequency of Checks for Understanding, the more information the teacher has about what students know and what they still need allowing for instruction to be targeted. However, the quality of typical Checks for Understanding may be more difficult to assess.

Checks for Understanding were defined as any skill-relevant teacher request or question requiring a verbal, written, or action student response. In cases where the teacher pressed for an elaborated verbal response, the Student Think-Aloud Metacognitive Strategy was scored instead. An unintended consequence of this procedure was that Verbal Checks for Understanding incidences were almost always representative of brief, unelaborated student responses. Arguably, these brief requests for information are the counterpart to Simple Feedback, a necessary, but low effect-size practice. Student Think-Aloud metacognitive strategy is a deeper and higher effect size strategy for checking student understanding. Student Think-Aloud, on average, accounted for less than 3% of intervals, indicating that only a small percentage of teacher Checks for Understanding demanded more than quick, possibly surface-level understanding.

In addition to frequent, less useful Checks for Understanding, a pattern in the type of Checks for Understanding most often used was evident. Verbal Checks for Understanding accounted for 81% of all Checks for Understanding. Verbal Checks for Understanding were most frequently one student at a time; rarely did teachers call for choral responses. Only 2.9% of intervals included

Written Checks for Understanding and even fewer Action Checks for Understanding (0.7%) were observed, suggesting a frequent unvaried pattern of question and verbal response. No teachers used white boards to check multiple student responses simultaneously. Also rarely did teachers use simple action responses such as thumbs up/thumbs down to quickly survey student understanding.

Marzano et al. (2001) characterized Checks for Understanding as “the heart of classroom practice,”(p.113). Yet teachers are found to be frequently unaware of the degree to which they question and cue students (Fillippone, 1998). Furthermore, consistent with study findings, teachers did not routinely check that all students were following a lesson and understand (A. D. Fisher & Frey, 2007; Marzano, 2007). A. D. Fisher and Frey (2007) stated, “knowing that six or seven students understand (i.e., those who raise their hands) is not the same as knowing that 32 do” (p.37). Although the impact of effective Formative Assessment is known to be great, a lack of conscious or purposeful application of EB practice is evident in the lack of informal assessment variety and depth observed.

Wiliam (2007) calculated that improving formative assessment practices in math could improve the US ranking into the top five countries of PISA. Marzano (2007) lamented, however, that teachers and teacher teams tend not to integrate such simple and powerful effective practices into their lesson plans, despite the availability of resources. Studies (e.g., Gleissman, Pugh, Dowden, & Hutchins, 1988; Redfield & Rousseau, 1981) show that teachers can dramatically improve their assessment skills when they become more conscious of automatic routines, learn about, and practice EB assessment techniques through training, coaching, and self-monitoring. Given the potential of Formative Assessment for enhancing student achievement, a focus on improving the depth, breadth, and consistency of assessment practices in Algebra 1 is warranted.

Intervention class designed for struggling math learners not different.

One of the principle questions of this study was to determine if instruction differed between single-period (1P) classes, geared for average math achievers, and double-period (2P) classes, geared to struggling math students. Results suggested that no significant teacher instructional practice differences were present. Similar percentages of EB Strategy Instruction, Elaborated Feedback, Formative Assessment, were observed. Other than double the time, differences in EB instruction were not detected.

The creation of separate classes for students with different levels of mathematics achievement was intended to allow teachers to target instruction to their specific needs. Considering research shows that struggling math students benefit most from highly explicit instruction, 2P teachers may benefit from training in these methods. These results should be shared with teachers and administrators to consider whether a focus on one or more explicit EB instructional practices, such as increasing the use of Explicit Modeling and

Review, Worked Examples, or Metacognitive Strategy Instruction, could benefit struggling math students in closing the achievement gap.

Implications for Practice

Results from this study suggest teachers were actively engaged in instruction. However, study results indicated that the most powerful EB instructional practices are limited in terms of frequency, variety, and depth of use with large differences between teachers. Several very highly effective instructional practices were used by only one or two teachers and/or for only a very small amount of class time. *Some* variability in instructional practices from one teacher or classroom to the next is not in and of itself disconcerting. Good teaching relies on the ability to respond to student needs based ongoing formative assessment, meaning no two classrooms will look *exactly* the same. However, *high* variability is likely to result in high variability of instructional outcomes (Rockoff, 2004; Sanders & Rivers, 1996; Weisberg et al., 2009). Teacher-to-teacher instructional practice differences result in avoidable inconsistencies in student achievement, shown in some studies to amount to years of learning (Sanders & Rivers, 1996).

Why are algebra classrooms so different in instructional practices? Differences may be attributed a number of factors, among them (a) poor pre-service teacher training in highly effective instructional practices, (b) ineffective professional development (PD) or poor implementation practices post-PD (e.g., lack of coaching, ongoing reinforcement) and (c) a lack of systematic and widespread sharing of expertise among teachers.

Variability in teacher practices is not surprising given the significant variability in the composition and quality of pre-service teacher training programs and certification requirements. In the U.S, "states' numerous licensing standards differ from one another and do not convey a clear, coherent vision of teaching and learning" (Darling-Hammond & Cobb, 1995, p. 229). In addition, pre-service training in mathematics is typically composed of separate coursework focused on content expertise and pedagogy, with little overlap, and differences in quality and emphasis (Darling-Hammond & Cobb, 1995; Kilpatrick et al., 2001). Darling-Hammond (1995) also describes "loose-linkage" between pre-service training and real-world experience evident in most teacher practica experiences arguing, "colleges have traditionally exerted little influence on the nature or quality of the practicum." Lessons can be learned from the practices in high-performing countries on the PISA. For example, in Finland, teacher preparation, on average, involves much more extensive coursework and fieldwork with a strong emphasis on "how to teach- using research-based state-of-the-art practices." (Organisation for Economic Cooperation and Development, 2011, p. 126).

Once in the classroom, variability of U.S. teacher professional development is the norm. Most school districts allocate as little as 1% of their budget to professional development, compared to 8-10% allocated in most businesses (Darling-Hammond & Cobb, 1995). Only a portion of this small amount of professional development is focused on pedagogy. Further problems are evident in the allocation of resources for implementation, such as follow-up training, fidelity checks, and coaching. Clancy (2006) reported that the Institute of Education Sciences spent 96% of its funding on developing *new* interventions and less than 4% on supports for their implementation. If EB highly effective instructional methods are to bridge the gap from research into practice, it is imperative to improve the effectiveness of both training and implementation supports (See Fixsen, Blase, Metz, & Van Dyke, 2013).

Again, lessons can be learned from high-performing countries on the PISA. In high-performing countries, teachers are developed into highly effective instructors through extensive professional development and coaching in the classroom setting. For example, in Singapore, teachers receive 100 hours of paid training per year; in Shanghai, teachers receive 240 hours every 5 years.

In addition to improving pre-service training and on-the-job professional development, increasing the frequency and quality of use of EB instructional practices can be accomplished through the collaborative efforts of teacher teams. Research on teacher leadership and professional development shows that effective teacher teams can significantly improve and increase individual teachers' use of EB practices, in turn, significantly enhancing student achievement (Lee & Smith, 1996; Moller, Mickelson, Stearns, Banerjee, & Bottia, 2013). Eaker, DuFour, and DuFour (2002) stated

> If there is anything that the research community agrees on, it is this: The right kind of continuous, structured teacher collaboration improves the quality of teaching and pays big, often immediate, dividends in student learning and professional morale in virtually any setting (p. xii).

Highly effective teacher teams have leaders who direct them to "ensure sound, ever-improving instruction and lessons…[and] discuss progress on common quarterly assessments" (Schmoker, 2011). Effective teams share and highlight effective practices using model classrooms and peer walk-throughs.

Effective collaboration is also seen in the highest performing countries on the PISA. Reportedly, these countries,

> generally consider teaching a profession where teachers work together to frame what they believe good practice to be, conduct field-based research to confirm or disprove the approaches they develop, and then judge their colleagues by the degree to which they use practices proven effective in their classrooms. This amounts to the collective search for ever more effective practices of the sort seen in Canada, Finland, Japan, Shanghai-China and Singapore (Organisation for Economic Cooperation and Development, 2011, p. 242).

Collaborative, data-driven teams of teachers consistently planning and evaluating their outcomes represent a continuous improvement system with the power to increase the implementation of EB practices leading to improved student outcomes.

Limitations /Threats to Validity

A number of limitations impact the degree to which the results can be generalized to algebra classes across the country. Most notably, threats to the conclusions drawn are based on (a) instrumentation, (b) sample size, and (c) representativeness.

Instrumentation. The OEBI developed for use in this study included a large number (16) of discrete EB instructional practices, aimed at discovering the degree to which these practices are being used in Algebra 1 classes. The OEBI was not difficult to use and resulted in high levels of inter-observer agreement (91%). Observational methods that record simply the presence or absence of a behavior are often highly reliable, as they are low-inference and result in high levels of accuracy (Millman & Darling-Hammond, 1989).

Although the OEBI may be highly reliable in determining the frequency and type of EB instruction typically occurring in Algebra 1 classrooms, systematic observation methods, in general, have limitations. Methodological concerns surface based on the use of a researcher-developed observation instrument related to the validity of the content and the process. First, the integrity of the OEBI has not been validated through replication studies.

Second, researcher-developed instruments such as the OEBI engender questions about construct validity, especially criterion-related validity, which is concerned with the accurate labeling of research operations (Shadish, Cook, &

Campbell, 2002). Shadish et al. (2002) state that we may "never be certain that anything is labeled with perfect accuracy" (p.468). The further one moves from discrete variables and their associated effect sizes to researcher-created categories, the chosen labels may become debatable. For example, for the purposes of analysis the discrete variables of Explicit Modeling, Explicit Review, Guided Practice and Independent Practice were combined to form the Direct Instruction category. Direct Instruction may in fact encompass several other discrete OEBI variables including the EB strategies, Mastery Goals and Student Think-Aloud, as well as Elaborated Feedback and Formative Assessment variables. A clear difficulty in observation studies is the tendency for different researchers to group variables very differently making comparison and confirmation of results problematic. Accurate labeling concerns represent both a methodological limitation and a practical concern.

The process of how variables are observed is also subject to limitations. When the OEBI was developed, observers were required to record all instructional practices occurring in a 30-second interval and as a result, the combined percentage of instructional practices exceeded 100%. Sums greater than 100% can be difficult to interpret. However, it was often the case that more than one instructional practice occurred within a 30-second interval; limiting recording procedures to one procedure per interval results in under-reporting of EB practices (Cornett, 2010). Simultaneous recording of instructional practices allows for a far more authentic representation of typical classroom instruction (L.

W. Anderson & Burns, 1989). In this study, on average 134 instructional practices were used in a 45-minute period (90, 30-second intervals). Using a non-simultaneous recording procedure would eliminate 44 of these practices from consideration.

However, some refinements in the procedures may improve the interpretability of results. For example, Independent Practice even when done for 85% of a class period with the teacher sitting at her desk checking emails was considered EB instruction. Because Physical Monitoring was marked only in whole-intervals (WIR) it is not clear in OEBI results what the teacher was in fact doing during Independent Practice. It could prove beneficial for teacher behavior during Independent Practice to be recorded, especially if the tool is to be useful as a formative evaluation tool for feedback purposes to improve teacher or program effectiveness.

In addition, because this study was focused on teacher behavior, small group student work represented a problem. The OEBI, unlike Cornett’s (2010) COS, did not record when instruction was whole-group or small group. Given small group work is prevalent in modern math classrooms, it seems it would be important to differentiate the teacher’s behavior during small group work versus whole group instruction. In future studies, the return to concomitant recording of learning arrangement (i.e., small group vs. large group) is recommended.

Sample Size. Other threats to this study's conclusions come from the sample size. Specifically, the analysis of variance between the two class types were based on a small number of classrooms (*n*=6) and teachers (*n*=15) reducing statistical power to detect differences and increasing the probability of a Type 2 error.

In addition, critics of observational research contend that the actual amount of time or number of observations needed to obtain a valid measure of instruction is not known (Waxman, Hilberg, & Tharp, 2004). Twelve classrooms total were each observed twice. Two observations per classroom may be considered an acceptable minimum, but 24 observations may not reflect all possible instructional configurations.

Reactivity. One of the primary methodological concerns of observational research is that the process of observation can be obtrusive interfering with the ability to draw valid conclusions (Waxman et al., 2004). Because teachers are aware that they are being observed and potentially anxious, they may engage in behaviors not typical to everyday practice. Observers can elicit a problem of teachers "faking good," or engaging in atypical practices meant to fulfill the perceived expectations of the observer. Reactivity is not likely to be a serious concern in this study because teachers noted that 94% of the observed instruction was *Very Typical* or *Somewhat Typical* on the Post-Observation Teacher Questionnaire (See Appendix X).

Representativeness. Because this study relied on data collected from two schools in the same community, findings may not be generalizable to other schools or communities (Shadish et al., 2002). Generalization of findings to other subjects or settings should be done with caution. However, because almost all of the Algebra 1 teachers (15 out of 17) from both schools participated and each classroom was observed two times, the results and study implications have local relevance.

Future Research

Results of this study were based on a small sample size from a single community; a small number of teachers and classrooms by type (1P and 2P classes) were included. It is recommended that this study be replicated with more algebra teachers and in different communities in order to determine if the current results are representative of a U.S. instructional pattern.

The OEBI may also be modified to assess the frequency of particular EB practices or the composition of complex subcategories, such as Practice. For example, Guided Practice was an instructional method that could contain other EB practices, such as Feedback and Checks for Understanding. In future research it may be useful to specifically investigate the typical composition of Guided Practice in terms of the frequency of other EB practices occurring simultaneously (e.g. frequency of Feedback and Checks for Understanding). Inclusion of achievement outcomes would also allow investigators to discover the most effective composition of high quality Guided Practice.

As a part of the current study, post-observation teachers were asked to label the primary instructional focus, i.e. *new material, review of previously learned material,* or *application*. It is recommended that future research in this area determine if the frequency of EB instructional practices differs based on the instructional focus of the lesson. Large instructional practice variability between teachers may be in part explained by differences in the lesson focus.

Summary

According to 40 years of research, three categories of instructional practices are consistently shown to enhance student achievement in math and across subject areas: (a) EB Strategy Instruction, (b) Feedback, and (c) Formative Assessment. It was the hypothesis of this study that Grade 9 Algebra 1 classrooms do not routinely use highly effective, evidence-based (EB) practices to enhance their instruction. Study results suggested that although high school Algebra I teachers were engaged in instruction more often than expected, the frequency of EB instructional practices varied remarkably among teachers. Furthermore, the preponderance of teaching time was spent in two forms of practice with little time devoted to other EB Strategy Instruction and informal Formative Assessment practices often lacked variety and depth. Last, the frequency or type of EB instructional practices used did not differ between classes designed for students with average math skills (1P) compared to classes designed for lower skilled students (2P). Given, the pervasive U.S. student lack of algebra readiness by high school and the established importance of Algebra 1 for students' future success, it is critical that Grade 9 Algebra 1 teachers provide high quality curriculum and instruction. And yet, many EB practices continue to fail to make it into the classroom. It is time, arguably imperative, for educators to use the available knowledge base to improve instruction in a purposeful, collaborative, and systematic manner in the United States. Recommended methods for increasing the use of highly effective EB instructional practices include: (a) system-wide improvements in pre-service teacher training in highly effective instructional practices, (b) more effective on-the-job professional development and implementation practices, and (c) the use of structured professional learning communities focused on improving pedagogy.

References

ACT. (2013). ACT Explore Technical Manual. from ACT,Inc.

Adams, G.L., & Englemann, S. (1996). *Research on direct instruction: 20 years beyond DISTAR*. Seattle, WA: Educational Achievement Systems.

Alvermann, D.E. (1981). The compensatory effect of graphic organizers on descriptive text. *Journal of Educational Research, 75*, 44-48.

Anderson, J.R. (1995). *Learning and Memory: An Integrated Approach*. New York: Wiley.

Anderson, L.W., & Burns, R.B. (1989). *Research in classrooms: The study of teachers, teaching, and instruction*. Oxford: Pergamon Press.

Archer, A.L., & Hughes, C.A. (2011). *Explicit Instruction: Effective and Efficient Teaching*. New York & London: The Guilford Press.

Artzt, A.F., & Armour-Thomas, E. (1992). Development of a cognitive-metacognitive framework for protocol analysis of mathematical problem solving in small groups. . *Cognition and Instruction, 9*(2), 137-175.

Ausubel, D.P. (1968). *Educational Psychology: A cognitive view*. New York: Holt.

Baker, S., Gersten, R., & Lee, D. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *Elementary School Journal, 10*, 51-73.

Bangert-Downs, R.L., Kulik, C.C., Kulik, J.A., & Morgan, M. (1991). The instructional effects of feedback in test-like events. *Review of Educational Research, 61*(2), 213-238.

Barbash, Shepard. (2011). Clear teaching: With direct instruction, Siegfried Engelmann discovered a better way of teaching: Education Consumers Foundation.

Berliner, D.C. (1984). The glass half-full: A review of research on teaching. In P. L. Hosford (Ed.), *Using What We Know About Teaching.* (pp. 51-84). Virginia: Association for Supervision and Curriculum Development.

Biddle, R. (2012). America's woeful public schools: TIMSS sheds light on the need for systematic reform from http://dropoutnation.net/2012/12/11/americas-woeful-public-schools-timms-sheds-light-on-the-need-for-systemic-reform/

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principals, policies, and practice, 5*(1), 7-74.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation & Accountability, 21*(1), 5-31. doi: 10.1007/s11092-008-9068-5

Booth, J.L., Lange, K.E., Koedinger, K.R., & Newton, K.J. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction, 25*, 24-34.

Bromley, K., Irwin DeVitis, L., & Modlo, M. (1999). *50 Graphic organizers for reading, writing and more.* . New York: Scholastic Professional Books.

Brophy, J.E., & Good, T.L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (3 ed., pp. 328-377). New York: MacMillan Publishing Company.

Burns, R.B. . (1984). How time is used in elementary schools: The activity structure of classrooms. In L. W. Anderson (Ed.), *Time and school learning: Theory, research and practice*. London: Croom Helm.

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

Carnine, D., & Gersten, R. (2000). The nature and roles of research in improving achievement in mathematics. *Journal for Research in Mathematics Education, 31*(2), 138-143.

Carnine, D., Jones, E., & Dixon, R. (1994). Mathematics: Education tools for diverse learners. *School Psychology Review, 23*(3), 406-427.

Carpenter, T., & Fennema, E. (1996). A longitudinal study of learning to use children’s thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27*(4), 403-434.

Carroll, W.M. (1992). *The use of worked examples in teaching algebra*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Carroll, W.M. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology, 86*, 360-367.

Clancy, C. . (2006). The $1.6 trillion question: If we're spending so much on healthcare, why so little improve- ment in quality? . *Medscape General Medicine, 8*(2), 58.

Cohen, J. (1988). *Statistical powe analysis for the behavioral sciences*. Hillsdale,NJ: L. Erlbaum Associates.

Cooper, G.A., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology, 79*, 347-362.

Cornett, J. (2010). *What's evidence got to do with it?: An observational study of research-based instructional behavior in high school classes* (M.S.Ed.), University of Kansas, Ann Arbor, MI. (1476574.)

Cybriwsky, C. A., & Schuster, J. W. . (1990). Using constant time delay procedures to teach multiplication facts. *Remedial and Spe-cial Education, 11*(1), 54-59.

Danielson, C. (2007). Enhancing professional practice: A framework for teaching: Association for Supervision & Curriculum Development.

Darling-Hammond, L., & Cobb, V.L. (1995). *Teacher preparation and professional development in APEC members: A comparative study*. Washington D. C. : Department of Education.

Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. . (2009). Professional learning in the learning profession. Washington, DC:: National Staff Development Council.

Dempster, F. N. (1991). Synthesis of research on reviews and tests. *Educational Leadership, 4, 7 1 - 7 6 . 4*, 71-76.

Deshler, D. (2003). Intervention research and bridging the gap between research and practice. *Learning Disabilities: A Contemporary Journal, 1*(1), 1-7.

Deshler, D., Schumaker, J.B., Lenz, B.K., Bulgren, J.A., Hock, M.F., Knight, J., & Ehren, B.J. (2001). Ensuring content-area learning by secondary students with learning disabilities. *Learning Disabilities Research and Practice, 16*(2), 96-108.

Donovan, J, J,, & Radosevich, D.J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology, 84*(5), 795-805.

Duzinski, G.A. (1987). *The educational utility of cognitive behavior modification strategies with children: A quantitative synthesis*. Dissertation. University of Chicago. Chicago,IL.

Eaker, R., DuFour, R., & DuFour, R. (2002). *Getting started: Reculturing schools to become professional learning communities*. Bloomington, IN: National Education Service.

Education, U.S. Department of (Producer). (2010). Education Secretary Arne Duncan issues statement on the results of the Program for International

Student Assessment. [Press Release] Retrieved from http://www.ed.gov/news/press-releases/education-secretary-arne-duncan-issues-statement-results-program-international-s

Engelmann, S., & Bereiter, C. . (1966). *Teaching disadvantaged children in the preschool*. Engelwood Cliffs, NJ: Prentice-Hall.

Fillippone, M. (1998). *Questioning at the elementary level.* (Master's), Kean University. ERIC Document Reproduction Service database.

Fisher, A.D., & Frey, N. (2007). *Checking for understanding*. Alexandria,VA: ASCD.

Fisher, R.A. (1934). *Statistical Methods for Research Workers. 5th ed.* Edinburgh: Oliver and Boyd.

Fixsen, D., Blase, K., Metz, A., & Van Dyke, M. (2013). Statewide implementation of evidence-based programs. *Council for Exceptional Children, 79*(2), 213-230.

Fuchs, L.S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children, 51*, 199-208.

Fuchs, L.S., Fuchs, D., & Hamlett, C.L. (1994). Strengthening the connection between assessment and instructional planning with expert systems. *Exceptional Children, 61*, 138-146.

Fyfe, Emily R., Rittle-Johnson, Bethany, & DeCaro, Marci S. (2012). The Effects of Feedback during Exploratory Mathematics Problem Solving: Prior Knowledge Matters. *Journal of Educational Psychology, 104*(4), 1094-1108.

Gage, N.L. (1984). What do we know about teaching effectiveness? . *Phi Delta Kappan, 66*(2), 87-93.

Gamoran, A. (1987). The stratification of high school learning opportunities. *Sociology of Education, 60*(July), 135-155.

Garofalo, J., & Lester, F. (1985). Metacognition, cognitive monitoring, and mathematical performance. *Journal for Research in Mathematics Education, 16*, 163-176.

Germann, G. (2010). Preface: Thinking of yellow brick roads, emerald cities, and wizards. In M. R. Shinn & H. Walker (Eds.), *Interventions for achievement and behavior problems in a three-tier model including RTI*. Bethesda,MD: NASP.

Gersten, R., Chard, D., Jayanthi, M., & Baker, S. (2006). Experimental and quasi-experimental research on instructional approaches for teaching mathematics to students with learning disabilities: A research synthesis. Signal Hill, CA: Center on Instruction/RG Research Group.

Gersten, R., Chard, D.J., Jayanthi, M., Baker, S.k., Morphy, P., and Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research, 79*(3), 1202-1243. doi: 10.3102/0034654309334431

Gersten, R., & Clarke, B.S. (2013). Effective strategies for teaching students with difficulties in mathematics. In J. Reed (Ed.). Reston, VA: National Council of the Teachers of Mathematics.

Gijbels, D., Dochy, F., Van den Bossche, P., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research, 75*(1), 27-61.

Givvin, K.B., Hiebert, J., Jacobs, J.K., Hollingsworth, H., & Gallimore, R. (2005). Are there national patterns of teaching? Evidence from the TIMSS 1999 video study. *Comparative Education Review, 49*(3), 311-343.

Gleissman, D.H., Pugh, R.C., Dowden, D.E., & Hutchins, T.F. (1988). Variables influencing the acquisition of generic teaching skill. *Review of Educational Research, 58*(1), 25-46.

Goldman, S.R. (1989). Strategy instruction in mathematics. *Learning Disabilities Quarterly, 12*, 43-55.

Good, T.L., Grouws, D.A., & Ebmeier, H. (1983). *Active mathematics teaching*. New York: Longman.

Grobe, Cornelia S., & Renkl, Alexander. (2007). Finding and Fixing Errors in Worked Examples: Can This Foster Learning Outcomes? *Learning and Instruction, 17*(6), 612-634.

Gump, P.V. (1967). The classroom behavior setting: Its nature and relation to student behavior (Final report). Washington, DC: U.S. Office of Education, Bureau of Research.

Hanushek, E.A., Peterson, P.E., & Woessmann, L. (2012). Achievement Growth: International and U.S> state trends in student performance: Harvard's Program on Educational Policy and Governance & Education Next.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research, 66*(2), 99-136.

Healy, J.M. (1990). *Endangered Minds: Why Our Children Don't Think.* New York: Simon & Schuster.

Hedges, L.V. (1987). How hard is hard science, how soft is soft science? The empirical cumulativeness of research. *American Psychologist, 42*, 443-455.

Heward, W.L. (2003). Ten faulty notions about teaching and learning that hinder the effectiveness of special education. *The Journal of Special Education, 36*(4), 186-205.

Hoffer, T.B., Rasinski, K.A., & Moore, W. (1995). Social background differences in high school mathematics and science coursetaking and achievement. Washington, D.C.: National Center for Education Statisitcs.

Holliday, B., G.J., Cuevas., Moore-Harris, B., Carter, J.A., Marks, D., Casey, R.M., . . . Hayek, L.M. (2005). *Algebra I*. Columbus, OH: McGraw-Hill Companies.

Jetter, A. . (1993). Mississippi Learning. (Cover Story). *New York Times Magazine, 142,* 28.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). Adding it up: Helping children learn mathematics: National Research Council.

Kirschner, Paul A., Sweller, John, & Clark, Richard E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist, 41*(2), 75-86. doi: 10.1207/s15326985ep4102_1

Kluger, Avraham N., & DeNisi, Angelo. (1998). Feedback Interventions: Toward the Understanding of a Double-Edged Sword. *Current Directions in Psychological Science (Wiley-Blackwell), 7*(3), 67-72. doi: 10.1111/1467-8721.ep10772989

Kroesbergen, E.H., & Van Luitt, J.E. . (2003). Mathematics interventions for children with special education needs: A meta-analysis. *Remedial and Special Education, 24*, 97-114.

Labuhn, Andju Sara, Zimmerman, Barry J., & Hasselhorn, Marcus. (2010). Enhancing Students' Self-Regulation and Mathematics Performance: The Influence of Feedback and Self-Evaluative Standards. *Metacognition and Learning, 5*(2), 173-194.

Ladson-Billings, G. (1997). <It Doesn't Add up- African American Students' Mathematics Achievement.pdf> *Journal for Research in Mathematics Education* (Vol. 28, pp. 697-708): National Council of the Teachers of Mathematics.

Lange, K. E., Booth, J. L., & Newton, K. J. (2014). Learning Algebra from Worked Examples. *Mathematics Teacher, 107*(7), 535-540.

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

Lee, V., & Smith, J.B. (1996). Collective responsibility for learning and Its effects on gains in achievement for early secondary school students. *American Journal of Education, 104*, 103-147.

Lenz, B.K., Alley, G.R., & Schumaker, J.B. (1987). Activating the inactive learner: Advance organizers in the secondary content classroom. *Learning Disabilities Quarterly, 10*, 53-67.

Lloyd, J.W., Forness, S.R., & Kavale.K.A. (1998). Some methods are more effective than others. *Intervention in School and Clinic, 33*(4), 195-200.

Loveless, T. (2013). The 2013 Brown Center report on American education: How well are American students learning? (Vol. 3): Brown Center on Education Policy at Brookings.

Loveless, T., Martin, M.O., Mullis, I.V.S., Schmidt, W. , & Kilpatrick, J. (2008, January, 2008). *Lessons learned: What international assessments tell us about math achievement*, Washington, D.C.

Ma, X. (2001). A longitudinal assessment of antecedent course work in mathematics and subsequent mathematical attainment. *The Journal of Educational Research, 94*(1), 16-28.

Marcucci, R.G. (1980). *A meta-analysis of research of teaching mathematical problem solving.* University of Iowa, Dissertation Abstracts International, 41, 06A.

Martella, R.C., Nelson, J.R., Morgan, R.L., & Marchand-Martella, R.L. (2013). *Understanding and Interpreting Educational Research*. New York, New York: The Guilford Press.

Marzano, R.J. (2007). *The art and science of teaching*. Alexandria, VA: ASCD.

Marzano, R.J. (2012). It's how you use a strategy. *Educational Leadership*(December 2011).

Marzano, R.J., Pickering, D.J., & Pollack, J.E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.

McKinseyandCompany. (2009). The economic impact of the achievement gap in America's schools.

McKinseyonSociety (Producer). (2010, July 12, 2014). Byron Auguste-How The achievement gap In U.S. schools affects the economy [Video webcast] Retrieved from http://www.youtube.com/watch?feature=player_embedded&v=hP_bX9IkVQk

Merkley, D.M., & Jefferies, D. (2001). Guidelines for implementing a graphic organizer. *Reading Teacher, 54*(4), 350-357.

Mevarech, Z.R., & Kramarski, B. (2003). The effects of metacognitive training versus worked out examples on students' mathematical reasoning. *British Journal on Educational Psychology, 73*, 449-471.

Millman, J., & Darling-Hammond, L. (1989). *The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers*: SAGE Publications.

Moller, S., Mickelson, R.A., Stearns, E., Banerjee, N., & Bottia, M.C. (2013). Collective pedagogical teacher culture and mathematics achievement: Differences by race, ethnicity, and socioeconomic status. *Sociology of Education, 86*(2), 174-194. doi: 10.1177/0038040712472911

Montague, M., & Applegate, B. (1993). Mathematical problem-solving characteristics of middle school students with learning disabilities. *Journal of Special Education, 27*, 251-261.

Montague, M., Enders, C., & Dietz, S. (2011). Effects of Cognitive Strategy Instruction on Math Problem Solving of Middle School Students with Learning Disabilities. *Learning Disability Quarterly, 34*(4), 262-272.

Moses, R. (1995). Algebra: The New Civil Right. In e. a. E. In C. Lacampagne (Ed.), *The Algebra Initiative Colloquium* (Vol. Volume II, pp. 53-67). Washington,D.C.: US Department of Education.

Mullis, I.V.S., Martin, M.O. , Foy, P. , & Arora, A. (2012). TIMMS 2011 international results in mathematics: International Association for the Evaluation of Educational Achievement (IEA).

National Center for Education Statistics. (2004). The condition of education 2004. Washington, D.C.: U.S. Department of Education.

National Center for Education Statistics. (2011). *The Nation's Report Card: Mathematics 2011 (NCES 2012–458).* . Washington,D.C.: U.S. Department of Education.

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

NCTM. (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: National Council of Teachers of Mathematics Commission on Standards for School Mathematics.

NMAP. (2006). *National Math Advisory Panel: Chapter 3 Report of the task group on conceptual knowledge and skills*. US Department of Education.

NMAP. (2008). The final report of the National Mathematics Advisory Panel: U.S. Department of Education.

Nuthall, G.A. (2005). The cultural myths and realities of classroom teaching and learning: A personal journey. *Teachers College Record, 107*(5), 895-934.

Ogden, L. (1992). Why aren't effective teaching practices adopted? *Journal of Applied Behavior Analysis, 25*(1), 21-26.

Organisation for Economic Cooperation and Development. (2004). Learning for tomorrow’s world: First results from PISA 2003. Paris.

Organisation for Economic Cooperation and Development. (2007). PISA 2006: Science competencies for tomorrow’s world (Vol. 1). Paris: OECD.

Organisation for Economic Cooperation and Development. (2011). Strong performers and successful reformers in education: Lessons from PISA for the United States.

Ostad, S. A., & Sorensen, P. M. (2007). Private speech and strategy-use patterns bidirectional comparisons of children with and without mathematical difficulties in a developmental perspective. *Journal of Learning Disabilities, 40*(1), 2-14.

Pianta, R.C., La Paro, K.M., & Hamre, B.K. (2007). *Classroom Assessment Scoring System – CLASS*. Baltimore, MD: Brookes.

Polya, G. (1957). *How to solve it: A new aspect of mathematical method*. London: Penguin.

Poplin, M.S. (1988). The reductionistic fallacy in learning disabilities replicating the past by reducing the present. Journal of learning disabilities. *Journal of Learning Disabilities, 21*(7), 389-400.

Przychodzin-Havis, A. M., Marchand-Martella, N. E., Martella, R. C. , & Azim, D. (2004). Direct Instruction mathematics programs: An overview and research summary. *Journal of Direct Instruction, 4*(1), 53-84.

Redfield, D.L., & Rousseau, E.W. (1981). A meta-analysis of experimental research on teacher questioning behavior. *Review of Educational Research, 51*(2), 237-245.

Robinson, D.H., & Kiewra, K.A. (1995). Visual argument: Graphic organizers are superior outlines in improving learning from text. *Journal of Educational Psychology, 87*, 455-467.

Rockoff, J.E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94*(2), 247-252.

Rosenshine, B., & Meister, C.C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research, 64*(4), 479-530.

Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (3 ed., pp. 379-391). New York: MacMIllan Publishing Company

Sanders, W.L., & Rivers, J. (1996). Cumulative and residual effects of teachers on future student academic achievement. In U. o. T. V.-A. R. a. A. Center (Ed.). Knoxville, TN.

Sawada, D., Piburn, M., Falconer, K. , Turley, J., Benford, R. , & Bloom, I. . (2000). Reformed Teaching Observation Protocol: Technical Report No. IN00-1. Tempe, AZ: Arizona State University. Tempe, AZ: Arizona State University.

Schiller, K. S., & Muller, C. (2003). Raising the bar and equity? Effects of state high school graduation requirements and accountability policies on students' mathematics course taking. *Educational Evaluation and Policy Studies, 25*(3), 299-318.

Schmidt, W. H., Rotberg, I. C., & Siegel, A. . (2003). Too little too late: American high schools in an international context. In B. p. i. e. policy (Ed.), (Vol. 6, pp. 253-307): Brookings Institute

Schoenfeld, A.H. (1985). *Mathematical problem solving*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schoenfeld, A.H. (1987). What's all the fuss about metacognition? In A. H. Schoenfeld (Ed.), *Cognitive Science and Mathematics Education* (pp. 189-215). Hillsdale,NJ: Lawrence Erlbaum Associates.

Schoenfeld, A.H. (1989). Teaching mathematical thinking and problem solving. In L. B. Resnick & L. E. Klopfer (Eds.), *Toward rethinking curriculum*. Alexandria, VA: Association for Supervision and Curriculum Development Yearbook.

Schoenfeld, A.H. (1992). Learning to think mathematically: problem solving, metacognition, and sense-making in mathematics. In D. A. Grouws (Ed.), *Handbook on research on mathematics teaching & learning*. New York: MacMillan Publishing Group.

Schumaker, J. B., Deshler, D.D., Bulgren, J. A., Davis, B., Lenz, B. K., & Grossen, B. (2002). Access of adolescents with disabilities to general education curriculum: Myth or reality? *Focus on Exceptional Children, 35*(3), 1-16.

Seligman, M.E.P. (1990). *Learned optimism*. New York: Alfred A. Knopf.

Senko, C., & Hulleman, C.S. (2013). The Role of Goal Attainment Expectancies in Achievement Goal Pursuit. *Journal of Educational Psychology, 105*(2), 504-521.

Sfard, Anna. (1991). On the Dual Nature of Mathematical Conceptions: Reflections on Processes and Objects as Different Sides of the Same Coin. *Educational Studies in Mathematics, 22*(1), 1-36. doi: 10.2307/3482237

Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth Cengage Learning.

Shinn, M.R. (1998). *Advanced applications of curriculum-based measurement*. New York: Guilford.

Siegler, Robert S., & Chen, Zhe. (2008). Differentiation and integration: guiding principles for analyzing cognitive change. *Developmental Science, 11*(4), 433-448. doi: 10.1111/j.1467-7687.2008.00689.x

Stainback, S.E., & Stainback, W.E. (1992). *Curriculum considerations in inclusive classrooms: Facilitating learning for all students*. New York: Paul H. Brookes Publishing.

Stigler, J.W., & Hiebert, J. (1997). Understanding and improving classroom mathematics instruction: an overview of the TIMSS video study. *Germany, Japan, and the United States, 79*, 14-21.

Sweller, J., & Cooper, G.A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*(1), 59-89.

Theokas, C. (2010). Shut out of the military. Washington,D.C.: Education Trust.

Urdan, T., Midgely, C., & Anderman, E.M. (1998). The role of classroom goal structure in students' use of self-handicapping strategies. *American Educational Research Journal, 35*(1), 101-122.

Valverde, G.A., & Schmidt, W. H. (2007). Refocusing U.S. math and science education: International comparisons of schooling hold important lessons for improving student achievement. http://www.issues.org/14.2/schmid.htm

Wainwright, C. L., Flick, L. B., & Morrell, P. D. (2003). Development of instruments for assessment of instructional practices in standards-based teaching *Journal of Mathematics and Science: Collaborative Explorations, 6*(1), 21-46.

Walberg, H.J. (1984). Synthesis of research on teaching. In M. C. Wittrock (Ed.), *Third Handbook of Research on Teaching* Washington, D.C.: American Educational Research Association.

Walkington, C., Arora, P., Ihorn, S. , Gordon, J. , Walker, M. , Abraham, L. , & Marder, M. . (2011). Development of the UTeach Observation Protocol: A classroom observation instrument to evaluate mathematics and science teachers from the UTeach Preparation Program (UTeach Technical Report 2011-01). . Austin, TX: University of Texas at Austin.

Waxman, H.C., Hilberg, R.S., & Tharp, R.G. (2004). Future directions for classroom observation research. In H. C. Waxman, R. G. Tharp & R. S.

Hilberg (Eds.), *Observational research in U.S. classrooms: New approaches for understanding cultural and linguistic diversity* (pp. 266-277). New York: Cambridge University Press.

Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunck, J., Palcisco, A., & Morgan, K. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness: New Teacher Project.

Wittrock, M.C. (1984). *Third Handbook of research on Teaching*. Washington, D.C.: American Educational Research Association.

Wu, H. (2001). How to prepare students for algebra. *American Educator, 25*(2), 10-17.

Zhu, X., & Simon, H. A. . (1987). Learning mathematics from examples and by doing. *Cognition and Instruction, 4*, 137-166.

Zimmerman, B.J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich & M. Zeidner, (pp. 13-39). (Eds.), *Handbook of self-regulation* (pp. 13-39). San Diego, CA Academic Press.

Appendix A

Basic Skills Required for Success in Algebra 1

Fluency With Whole Numbers	By the end of elementary school, students should have an understanding of place value, and the ability to compose and decompose whole numbers, including a grasp of addition, subtraction, multiplication, and division, with the ability to apply the operations to problem solving. Computational facility rests on the automatic recall of addition and related subtraction facts, and of multiplication and related division facts. It requires fluency with addition, subtraction, multiplication, and division. A strong sense of number also includes the ability to estimate the results of computations.
Fluency with Fractions	Before they begin algebra course work, middle school students should have a thorough understanding of positive as well as negative fractions. They should be able to locate both positive and negative fractions on the number line; represent and compare fractions, decimals, and related percents; and estimate their size. They need to know that sums, differences, products, and quotients (with nonzero denominators) of fractions are fractions, and they need to be able to carry out these operations confidently and efficiently. They should encounter fractions in problems in the many contexts in which they arise naturally, for example, to describe rates, proportionality, and probability. The subject of fractions, when properly taught, introduces students to the use of symbolic notation and the concept of generality, both being an integral part of Algebra (Wu, 2001).
Particular Aspects of Geometry and Measurement	Sound treatments of the slope of a straight line and of linear functions depend logically on the properties of similar triangles. Furthermore, students should be able to analyze the properties of two- and three-dimensional shapes using formulas to determine perimeter, area, volume, and surface area. They should also be able to find unknown lengths, angles, and areas.

Table X. Adapted from NMAP (2006, Chapter 3, p. 41) conclusions regarding the critical foundations for success in algebra.

cusd300 10/2/2014 10:27 AM
Formatted: Heading 1, Space After: 0 pt, Widow/Orphan control, Adjust space between Latin and Asian text, Adjust space between Asian text and numbers, Tabs:Not at 0.5"

cusd300 10/2/2014 10:27 AM
Formatted Table

cusd300 10/2/2014 10:27 AM
Formatted: Line spacing: single

cusd300 10/2/2014 10:28 AM
Formatted: Line spacing: single

cusd300 10/2/2014 10:29 AM
Formatted: Space After: 12 pt, Line spacing: single, No widow/orphan control, Don't adjust space between Latin and Asian text, Don't adjust space between

cusd300 10/2/2014 10:29 AM
Formatted: Font:(Default) Book Antiqua

Appendix B
Participant Consent Form for Observations

Title of Study: Evidence-Based Algebra Instruction

The purpose of the study: I understand that the purpose of this study is to understand more about evidence-based instruction. This is not an experiment. The researcher will not attempt to change the manner in which this class is taught. I agree to the following during the 2013-14 school year.

1. The researcher may request to speak with me about my instruction post-observation, specifically the degree to which the instruction observed was typical or atypical.
2. I understand that:
   (a) Participation is strictly voluntary. I can refuse to answer any questions that I do not wish to answer.
   (b) The information gathered will not affect current or future teacher evaluations.
   (c) The information gathered will be confidential. Teacher and student names or any other identifying factors will be removed from any report or publication of the data or results.
   (d) I may opt out of the project at any time and for any reason I deem necessary with no repercussions if I give written notice to the researcher.
   (e) Declining participation in this study will not cause adverse actions to be taken against me.
   (f) I may benefit from this study as the general observation results and data analysis along with my personal data will be shared with me after the conclusion of the study.
   (f) The researcher will observe some class sessions during the semester but will not audio or videotape the classes.

I understand that this research study has been reviewed and approved by the CUSD 300 Office of Assessment and Accountability. For research-related problems or questions regarding subjects' rights, I can contact the Assistant Superintendent Secondary Level.

I have read and understand the explanation provided to me. I have had all my questions answered to my satisfaction, and I voluntarily agree to participate in this study. I have been given a copy of this consent form.

Teacher's Signature ______________________________ Date__________________

If I do NOT wish to participate, I will not return this form. No adverse actions will be taken against me if I choose this option.

Researcher's Signature ___________________________ Date __________________

cusd300 10/2/2014 10:30 AM
Formatted: Heading 1, Centered, Line spacing: single

cusd300 10/2/2014 10:32 AM
Formatted: Space After: 12 pt, No widow/orphan control, Don't adjust space between Latin and Asian text, Don't adjust space between Asian text and numbers

cusd300 10/2/2014 10:32 AM
Formatted: Font:(Default) Book Antiqua, 11 pt

Appendix C.

Study Title: Evidence-Based Algebra Instruction

cusd300 10/2/2014 10:33 AM
Formatted: Heading 1 Char

cusd300 10/2/2014 10:33 AM
Formatted: Centered, Space After: 10 pt, Line spacing: double, Widow/Orphan control, Adjust space between Latin and Asian text, Adjust space between Asian

cusd300 10/2/2014 10:33 AM
Formatted: Font:12 pt

Dear ___,

My name is Piper Stratton. I am a doctoral candidate in the School Psychology Department at National-Louis University. I am conducting a research study as part of the requirements of my Ed.D. in School Psychology, and I would like to invite you to participate.

I am studying evidence-based instructional practices in Grade 9 Algebra 1 classes. If you decide to participate, you will be asked to allow one or two researchers to observe in your Algebra 1 classes two times. The observations will take place at a mutually agreed upon time, and should last the duration of the class period(s).

Although you probably won't benefit directly from participating in this study, we hope that others in the community/society in general will benefit from increased knowledge regarding current typical instructional practices in Algebra 1 classes.

Participation is confidential. Study information will be kept in a secure location at National-Louis University. The results of the study may be published or presented at professional meetings, but your identity will not be revealed.

Taking part in the study is your decision. You do not have to be in this study if you do not want to. You may also quit being in the study at any time or decide not to answer any question you are not comfortable answering. Participation, non-participation or withdrawal will not affect your evaluations in any way.

We will be happy to answer any questions you have about the study. You may contact me at or my faculty advisor, (Mark Shinn, 847-275-7200, and markshinn@icloud.com) if you have study related questions or problems.

Thank you for your consideration. If you would like to participate, please open and sign the attached consent form and return it to Piper Stratton. I will also call you within the next week to see whether you are willing to participate.

With kind regards,

____

Piper Stratton