Investing in our Teachers: What Focus of Professional Development Leads to the Highest Student Gains in Mathematics Achievement?

Alejandra Salinas
University of Miami, AleSalinas9@aol.com

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_dissertations

Recommended Citation
https://scholarlyrepository.miami.edu/oa_dissertations/393
INVESTING IN OUR TEACHERS: WHAT FOCUS OF PROFESSIONAL
DEVELOPMENT LEADS TO THE HIGHEST STUDENT GAINS IN
MATHEMATICS ACHIEVEMENT?

By

Alejandra Salinas

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

May 2010
INVESTING IN OUR TEACHERS: WHAT FOCUS OF PROFESSIONAL DEVELOPMENT LEADS TO THE HIGHEST STUDENT GAINS IN MATHEMATICS ACHIEVEMENT?

Alejandra Salinas

Approved:

Walter Secada, Ph.D.
Professor of Education

Terri A. Scandura, Ph.D.
Dean of the Graduate School

Okhee Lee, Ph.D.
Professor of Education

Batya Elbaum, Ph.D.
Professor of Education

Soyeon Ahn, Ph.D.
Professor of Education
The purpose of this meta-analysis was to better understand the relationship between the substantive-content focus of professional development for inservice teachers of mathematics and their students’ achievement. That professional development for teachers enhances student achievement has been well established by recent studies; however, those meta-analyses have studied structural characteristics such as the duration and the format/delivery method of the professional development. It is important to understand how the focus of professional development relates to student achievement because different foci must still compete not only among themselves but also with other instructional-improvement strategies (such as high-stakes testing, accountability, and curriculum reform) for limited resources, such as time and money. Hence, having evidence that professional development works and, more importantly, a better understanding of what focus comprises more effective professional development is not just of theoretical importance, it is also a policy-relevant imperative. The study’s results indicated that the focus of professional development is, in fact, a significant and educationally important predictor of variation in student-achievement effect sizes.
### TABLE OF CONTENTS

| LIST OF FIGURES | iv |
| LIST OF TABLES | v |

**Chapter**

1. **INTRODUCTION AND PURPOSE** ............................................................. 1  
   Making the Best Investment in Professional Development ..................... 2  
   Structural Characteristics versus the Substantive Content of  
   Professional Development ............................................................... 7  
   Narrative-based Reviews of Substantive  
   Professional Development Characteristics .......................................... 10  

2. **DETERMINING THE SUBSTANTIVE FOCI OF  
   PROFESSIONAL DEVELOPMENT** ............................................................ 13  
   Theoretical Support for the Categorization Scheme ................................ 15  
   Empirical Evidence in Support of the Categorization Scheme ................ 24  

3. **DATA COLLECTION AND ANALYSIS** .................................................... 30  
   Searching the Literature ....................................................................... 30  
   Gathering Information from the Studies ............................................... 35  
   Data Analysis ..................................................................................... 45  
   The Problem with Multiple Foci per Study ......................................... 57  

4. **RESULTS** .......................................................................................... 59  
   Effect Sizes ....................................................................................... 59  
   RQ1: Does the Focus of Professional Development Predict Some  
   Variation of Effect Sizes? ................................................................. 66  
   RQ2: Which Focus Leads to the Greatest Student Gains? ..................... 68  

5. **CONCLUSIONS AND DISCUSSIONS** .................................................... 69  
   RQ1: Conclusions and Discussions ...................................................... 71  
   RQ2: Conclusions and Discussions ...................................................... 73  

6. **IMPLICATIONS AND FUTURE RESEARCH** ........................................... 82  

REFERENCES ......................................................................................... 85  
APPENDIX A: Coding Sheet .................................................................. 97  
APPENDIX B: Effect Size Data Table ...................................................... 118
LIST OF FIGURES

FIGURE 1: Theoretical model of outcomes of professional development ............ 15
FIGURE 2: Theoretical framework connecting changes in pedagogical knowledge
to changes in pedagogical decisions .................................................... 18
FIGURE 3: Theoretical framework connecting changes in beliefs to changes
in teacher practices ........................................................................... 21
FIGURE 4: Flowcharts of the literature search process .................................... 32
FIGURE 5: Effect sizes with 95% confidence intervals .................................... 65
FIGURE 6: Overall means and 95% CI for each focus ..................................... 67
FIGURE 7: Funnel plot demonstrating the “file-drawer” problem ....................... 70
LIST OF TABLES

TABLE 1: List of identified studies and key study characteristics ......................... 40
TABLE 2: Measures of student math achievement ............................................... 42
TABLE 3: Effect sizes ........................................................................................... 62
TABLE 4: Summary of results ............................................................................... 66
TABLE 5: Post-hoc contrasts ................................................................................ 68
CHAPTER 1
INTRODUCTION AND PURPOSE

The purpose of this study was to better understand the relationship between the focus of professional development and student achievement by determining, specifically:

1. Whether or not the focus of professional development significantly predicts some of the variation of effects of professional development on student gains in mathematics; and

2. If so, which focus of professional development leads to the highest student gains in mathematics?

Professional development interventions vary in their impacts on student achievement gains in mathematics; some interventions work better than others (Blank & de las Alas, 2009, Scher & O’Reilly, 2009; Slavin & Lake, 2008). Thus, the answers to these questions have the potential of impacting policy, practice, and research in ways that may increase teacher quality and, consequently, student achievement. For example, as increased attention is being given to the relatively low-achievement of America’s children in mathematics, the results of this study could help inform policy-makers and practitioners on how to best invest their resources for professional development in order to have the greatest effect on student learning. Moreover, this study attempted to fill a gap in the professional development literature; whereas most of the quantitative research has emphasized how the structural components of professional development cause variations in student achievement, this study sought to empirically investigate the impact of professional development’s substantive content on achievement. In order to do this, this study brought to bear quantitative research methods on narrative-based reviews of the research on professional development.
**Making the Best Investment in Professional Development**

Increasing the quality of teaching in American classrooms is one way to increase student achievement in mathematics. Value-added models have indicated that teachers are responsible for a large amount of variation in student achievement (McCaffrey, Lockwood, Koretz, & Hamilton, 2003). In other words, the quality of the teacher assigned to each of two otherwise-identical (or at least, very similar) students will significantly affect how much each of those students learns and subsequently performs in mathematics. According to value-added models, the differences in student gains in mathematics based on teacher quality has been found to range between 3 and 54 percentile points. Thus, increasing teacher quality has the potential to have large impacts on student gains. Professional development is one of the most common vehicles for increasing in-service teacher quality (Sowder, 2007).

There are additional reasons to support professional development for enhancing teacher quality. Teachers are being expected to adjust their classroom practices and to teach using new methods and techniques (Darling-Hammond & McLaughlin, 1995). Sometimes, teachers were trained prior to the reform movement and they need to adjust to frequently changing standards; they may not have learned the skills or knowledge necessary to teach based on the newly emerging standards for math education (Cook, 1997). In other instances, new research on education may emerge which requires teachers to update their understandings of learning theories, the nature of math, and what comprises good pedagogy (Coburn, 2003). In general, professional development programs are designed to change teachers’ practices, their attitudes, and/or their beliefs,
all of which are thought to lead to improved learning opportunities for their students (Guskey, 2002).

In addition to updating teacher knowledge and skills, low student achievement provides another (possibly more powerful) motivation for providing teacher professional development. According to the National Assessment of Educational Progress (NAEP), fewer than 40% of fourth graders scored at or above proficient in math on the Nation’s Report Card in 2009 (U.S. Department of Education, 2010). Despite the fact that this is three times as many fourth graders as achieved this level than was reported in 1990, the latest results are still disappointing because over 60% of American fourth graders still have been unable to demonstrate solid academic performance in math. Improving student achievement relies on improving teacher quality, which in turn relies on effective professional development (Higgins & Parsons, 2009). NCLB demands that each state make high-quality professional development available. Likewise, The Teaching Commission released a report entitled *Teaching at Risk: A Call to Action* (2004) that emphasizes “ongoing and targeted professional development” (p. 11) as necessary in order to guarantee America’s future.

*Elementary School Teachers*

Elementary school teachers need professional development in mathematics because, most commonly, they have been trained as generalists. This means that elementary teachers may have been trained to teach all core subjects, such as reading, science, social studies and math; but they have not developed specialized skills in the teaching of any one subject (Cuff, 1993; Greenberg & Walsh, 2008). The consensus is
that this type of preparation leaves elementary school teachers noticeably under-prepared to teach math (Greenberg & Walsh, 2008). Not only do the pre-service teachers not gain the mathematical content knowledge that is necessary to teach effectively; but also, they do not gain the self-confidence necessary to teach it effectively (Cuff, 1993). Lack of self-confidence and content knowledge leads elementary teachers to teach math mechanically as a series of algorithms instead of teaching for understanding (Cohen & Leung, 2004; Foss & Kleinsasser, 1996; Hill, 1997).

On average, elementary school teachers have been found to dislike math and suffer from math anxiety (Cohen & Leung, 2004; Foss & Kleinsasser, 1996). Thus, it is possible that these teachers will pass along their anxiety and dislike of math to their students (Buhlman & Young, 1982). Yet, somehow teachers are expected to overcome these obstacles and lay the mathematical ground-work for their students; ground-work that is essential to understanding math in future courses.

What are the logical conclusions of the above? First, there is a clear need to improve teacher preparation programs for teaching mathematics. Second, and more relevant to this study, elementary school teachers need professional development whose content focus will help them gain the content knowledge, skills, and confidence that they did not gain in their pre-service program in order to improve their quality of math teaching. Absent a specific content-focus, it seems unlikely professional development, regardless its structural characteristics, will help elementary school teachers develop the in-depth skills required to teach mathematics well.
Limited Resources for Professional Development

Unfortunately, professional development requires two very limited resources. The first resource necessary for professional development is money. According to Killeen, Monk, and Plecki (2002), school districts across the United States spent an average of $200 per pupil a year for professional development between 1992 and 1998. This equates to approximately 3% of total expenses. However, Odden, Archibald, Fermanich, and Gallagher (2002) argued that the manner in which this figure was computed fails to take into account several costs related to professional development. That number was calculated on an overall sum under the general category of instructional support. This creates two complications. First, not all the monies spent on instructional support necessarily go to professional development. Second, monies spent on materials and substitute teachers, for example, may not fall under instructional support despite the fact that they are necessary expenses for professional development. Odden, Archibald, Fermanich, and Gallagher (2002) describe how professional development might actually be much more expensive than studies such as those by Killeen, Monk, and Plecki (2002) lead their readers to believe.

As an example, the UCLA Math Content Program for Teachers (n.d.) estimated the cost of professional development. They assumed that a substitute teacher costs $150/day. Their least expensive professional development program costs $500/day. Thus, one teacher for one day of this program costs the district $650. If a school district wants to send 30 of its teachers to this professional development, the school district must pay $19,500. On the other hand, effective professional development is a long-term process that usually requires more than one day (Banilower, Boyd, Pasley, & Weiss, 2006). If
teachers attend 12 days of the UCLA professional development intervention, least expensive professional development program will cost the school district $243,000. In other words, $8,100 is needed per teacher; this is far more that the $200 that is allocated per teacher as reported by Killeen, Monk, and Plecki (2002). Effective professional development is expensive. For districts to get the best value for the money spent on professional development, they should be informed as to what type of professional development is most likely to have the largest impacts on student achievement.

The other limited resource is time. Effective professional development is thought to require a minimum of 100 hours of training over the course of the program (Banilower, Boyd, Pasley, & Weiss, 2006). Teachers struggle to meet the daily demands of teaching, including planning their lessons, grading student work, and collaborating with their colleagues (McDiarmid, 1995). For teachers, a day in professional development does not mean less work. On the contrary, having to create lesson plans for the substitute teachers only adds to the teacher’s work load. Time is a precious resource. If teachers are going to invest their time in professional development, it is in their best interest to invest the time on the most effective professional development program possible. Therefore, to make the best investments of money and time, districts need to be informed on what characteristics make professional development the most effective. As such, the primary purpose of this study was to determine which types of professional development would lead to the highest gains in students’ math achievement in order to inform policy-makers and practitioners about how to best invest their resources.
Structural Characteristics versus the Substantive Content of Professional Development

When the new standards movement arose, researchers and policy-makers alike realized that the reform affected the teachers in addition to the students. Changing curricula and testing alone did not seem to lead to sufficient student gains; teachers needed to be able to understand and teach these new standards in order to successfully raise students’ math achievement. Consequently, the standards reform led to increased interest in teacher education and professional development (Wilson & Berne, 1999).

Structural Characteristics of Professional Development

Prior to this new emphasis on teacher learning, professional development tended to be comprised of short workshops in which teachers were most often expected to passively gain knowledge while an expert lectured in front of the room (Putnam & Borko, 1997; Wilson & Berne, 1999). However among researchers studying teacher learning in this new era of reform, Abdal-Haqq (1995), Ball (1996), and Putnam and Borko (1997) proposed that professional development itself should be reformed so as to lead to enhanced changes in teacher knowledge and beliefs. As opposed to having teachers listen passively, Putnam and Borko argued that teachers should be encouraged to be active learners, teacher educators should model the pedagogy which they expect the teachers to use with their students, and teachers should be viewed as professionals. This list highlights some characteristics of how professional development should be conducted, but not what the content of the professional development should be (Wilson & Berne, 1999).
Porter, Garet, Desimone, and Birman (2003) studied aspects of professional development that effectively change teacher knowledge and beliefs. Professional development providers should expect active learning from teachers; professional development should target some specific content (i.e., professional development for teaching math, or teaching reading, etc., as opposed to professional development for teachers in general); and, professional development should align with other learning activities. Porter et al.’s (2003) results indicated that these three aspects, in combination with characteristics such as the duration of the professional development (longer is better) and whether or not there are opportunities for collective participation of teachers (collective participation is better), led to more positive changes in teachers’ beliefs, knowledge, and practices when compared to other characteristics of professional development. These characteristics also highlight how professional development should be conducted and not what it should be targeting.

Two recent meta-analyses have each specifically investigated the effects of professional development on student gains in mathematics (Blank & de las Alas, 2009, and Scher & O’Reilly, 2009). Both studies explored the relationship between the structural components of professional development and student achievement; and both replicated Porter et al.’s (2003) results as to which structural components are most effective.

However, effective professional development should consider what teachers should learn as well as how best to accomplish this goal (Wilson & Berne, 1999). If what is being taught is not meaningful, then how it should be taught is inconsequential at best. The structural characteristics discussed by Putnam and Borko (1997) and by Porter et al.
(2003) may be necessary for implementing successful professional development; yet, the substantive content of the professional development is what gives meaning to what teachers are learning and this content should, theoretically, differentially affect student learning.

Substantive Content of Professional Development

As opposed to the emphasis that has been placed on finding the structural features of professional development that lead to student gains, the empirical research attempting to find the effective substantive content of professional development is sparse. One study that looked at the effectiveness of professional development through its substantive content was Scher and O’Reilly (2009). In addition to using structural components of professional development to explain differences in student gains, Scher and O’Reilly investigated whether professional development focused on content or pedagogy had greater effects on students. They were unable to parse out differences between these foci and concluded that professional development focused on both simultaneously had a greater effect than either individually.

Slavin and Lake (2008) conducted a best evidence synthesis on the differential effects of three different types of math reforms aimed at increasing elementary student achievement: curricular changes, computer-aided instruction programs, and instructional process programs. Some of the instructional process programs contained a large professional development component which focused on developing teachers’ pedagogical skills. Their findings indicated that of the three different types of reform, those involving instructional process programs had the greatest impact on student achievement. They then
divided the instructional process programs into seven categories of differing approaches to changing teacher knowledge and practices. These seven categories focused on: a) cooperative learning, b) cooperative/individualized programs, c) direct instruction, d) mastery learning, e) math content, f) classroom practices and motivation, and g) supplemental programs. Slavin and Lake’s results suggested that professional development which alters how teachers and students interact (such as by creating cooperative learning groups) had the greatest positive effects on student achievement.

The above studies suggest that research is needed to identify and quantitatively test (a) whether the substantive content of professional development explains variations in the impact of professional development on student achievement and (b) what content professional development should focus on in order to have the greatest impacts on students’ math achievement. This research seems like the logical next step in discovering what would make the best investment for America’s teachers. Yet, there is a gap in the literature concerning this topic; only a few studies, such as the ones above, have investigated effective professional development through a substantive lens rather than a structural one. Therefore, this study sought to identify and empirically test the groupings of substantive content of professional development suggested by the literature and, consequently, fill this existing gap in the literature.

**Narrative-based Reviews of Substantive Professional Development Characteristics**

Although few quantitative studies have investigated which substantive foci of professional development effectively lead to greater gains in students’ math achievement, several articles suggest that one or another focus on teacher learning may lead to
increased math achievement. Fennema and Franke (1992) and Sowder (2007) conducted narrative-based reviews of that literature.

Fennema and Franke (1992) suggested that four components of teacher knowledge have received the majority of researchers’ attention: a) content knowledge, b) knowledge of learning, c) knowledge of mathematical representations, and d) pedagogical knowledge. Pedagogical knowledge in Fennema and Franke’s review is very broad and includes non-content oriented concepts such as classroom management, planning their lessons, and grouping students. They argue that teachers must learn these four components in order to teach math effectively; and thus, professional development should be focused on these components.

Much more recently, Sowder (2007) identified six categories of needs by teachers that should be met by professional development programs: a) developing a shared vision, b) increasing content knowledge, c) learning about how students learn, d) learning pedagogical content knowledge, e) learning about the role of equity, and f) increasing the sense of self as a math teacher. Professional development focused on one or more of these six categories has been found to effectively improve teaching quality through some combination of qualitative and quantitative studies. Thus, the development of teacher knowledge in one or more of these six categories should lead to student achievement in mathematics.

Some of Sowder’s (2007) categories these are re-workings of Fennema and Franke’s (1992) model. For example, “learning how students learn” and “knowledge of mathematics representations” are both aspects of “pedagogical content knowledge”.
However, the inclusion of creating a shared vision, equity, and sense of self reflects concepts that researchers have taken a greater focus on since 1992.

These two narrative reviews suggest that there are specific types of knowledge that teachers need in order to teach math effectively. In-service teachers should be helped to gain or add to their knowledge of these topics through professional development. However, this categorization schema of the knowledge that teachers should learn through professional development has not been tested empirically. Thus, the final purpose of this study was to empirically test the efficacy of the schema suggested by these reviews to positively affect students’ math learning.
CHAPTER 2
DETERMINING THE SUBSTANTIVE FOCI OF PROFESSIONAL DEVELOPMENT

The purpose of this study was to investigate the relationship between student gains in mathematics and the focus of the substantive content of professional development. The literature reviewed in Chapter 1, above, has defined the substantive content that the field of mathematics education has deemed important for professional development to focus on. That literature guided how this study initially defined the foci of professional development.

Although Scher and O’Reilly (2009) investigated some of the effects of the substantive content of professional development, their categories (content versus pedagogy) were very broad and encompass many types of professional development interventions. It is possible that variation of effects on student achievement within their categories caused their inability to differentiate between the effects of their categories. For example, professional development focused on pedagogy could include both direct instruction and cooperative learning. If either focus of professional development (i.e., direct instruction versus cooperative learning) resulted in larger effects than the other, the range of effects of professional development focused on pedagogy would be much wider than had the study used the smaller grain size of professional development focused on cooperative learning versus direct instruction.

Slavin and Lake’s (2008) categorization of the instructional process programs did use a smaller grain size. However, their schema was a post-hoc amendment to the study. Their categorization schema was based on an analysis of the studies included in the synthesis and not on an *a priori* theoretical framework; they provided no rationale for
their categories. This is not to say that their results are not valuable; on the contrary, their results help to inform educational decision makers and professional development designers about aspects of professional development programs that are worthwhile. However, insofar as math education professional development flows out of some substantially different theories about teaching and learning, those theoretical positions should be considered when creating a framework for an *a priori* categorization of professional development foci. The math education literature, as consolidated by Fennema and Franke (1992) and Sowder (2007), suggests groupings that differ from the schema used by Slavin and Lake.

Although Slavin and Lake’s categorizations were not based on an *a priori* classification scheme, some of their categories, such as cooperative learning, were found to produce significant effects in student learning. Moreover, some of Slavin and Lake’s categories were identified as important by Fennema and Franke’s (1992) model. Knowledge of cooperative learning is an example of what Fennema and Franke refer to as pedagogical knowledge. Additionally, many professional development interventions continue to focus on cooperative learning or classroom management. Despite not being included in Sowder’s model, Slavin and Lake’s (2008) results suggest that cooperative learning and classroom management are likely to have impacts on students’ math achievement. Therefore, this study adopted Sowder’s model of foci and added classroom management and cooperative learning. This study used the following categories to answer its research questions:

a) subject matter knowledge (SMK)

b) pedagogical content knowledge (PCK)
c) equity

d) creating a shared vision

e) increasing sense of self as a math teacher

f) cooperative learning strategies

g) classroom management

**Theoretical Support for the Categorization Scheme**

According to several theoretical models, professional development (moderated by various quality measures) should lead to some proximal outcomes, such as changes in teacher knowledge and beliefs. These, in turn, should lead to some intermediate outcomes, such as changes in teacher practices. Finally, changes in teacher practices should lead to changes in student achievement (Blank & de las Alas, 2009; Scher & O’Reilly, 2009). The current study also assumes this theoretical model to explain the effects of professional development. However, it adds that the focus of professional development might also moderate the effects on the outcomes. This adjusted theoretical model is depicted in Figure 1.

**Figure 1:** Theoretical model of outcomes of Professional Development
**Theory Underlying Content Knowledge & Pedagogical Content Knowledge**

According to Shulman (1986), Ball, Thames, and Phelps (2008) and Hill, Rowan, and Ball (2005), teachers’ knowledge can be broken into two general categories: subject matter knowledge and pedagogical content knowledge. Each of these, in turn, is often broken down as well. The breakdown of these types of teacher knowledge differs based on the researcher and the theoretical model being used. Ball, Thames, & Phelps (2008) separated subject matter content into common content knowledge, specialized content knowledge, and horizon knowledge; and they separated pedagogical content knowledge into knowledge of content and students, knowledge of content and teaching, and knowledge of content and curriculum. Regardless of the form of categorization, each model agrees that teacher knowledge for teaching math includes some combination of knowing math, knowing how to teach math, and knowing how students learn math.

Theoretical and logical arguments, as well as empirical results, point towards the necessity of having these types of knowledge in order to teach effectively. The argument is that in order for math teachers to teach students effectively, they need to start with what their students know, what their students’ concepts and misconceptions are, and how to build on these. Hence, teachers need to understand the math well enough to interpret what their students are saying, why certain misconceptions might exist, how the math that they wish their students to learn is organized, and how to challenge those misconceptions with a variety of examples and representations.

More broadly, knowledge about mathematical content, teaching mathematics, and learning mathematics enables math teachers to select important activities, explain concepts in many ways, ask helpful questions, and assess student understanding (Hill &
Lubienski, 2007; Ball & McDiarmid, n.d.). In other words, teaching mathematics is necessarily mathematical in nature; and thus, teaching mathematics demands knowledge of mathematics and how to teach mathematics (Ball, Thames, & Phelps, 2008). Thus, increasing teachers’ knowledge seems to be a vital component for increasing student achievement because the practices that can be changed through this knowledge (e.g.: selecting more valuable activities, using a wider range of representations, etc.) have been directly linking to high quality teaching.

Theory Underlying Cooperative Learning and Classroom Management

As opposed to pedagogical content knowledge (PCK), pedagogical knowledge is thought of as unrelated to the content being taught. Pedagogical knowledge refers to basic skills, strategies, and behaviors that all teachers should know and use. Thus, knowledge of cooperative learning and classroom management are forms of pedagogical knowledge. From both a theoretical and empirical standpoint, the development of effective teacher behavior is a necessary component in attempts to improve student achievement (Brophy, 1986). Following the model in Figure 1 above, professional development focused on these types of pedagogical strategies can lead to improvements in teachers’ knowledge of pedagogy, which leads to improvements in practices, and ultimately improvements in student achievement.

The connections between professional development focused on cooperative learning and classroom management to changes in teacher knowledge of these pedagogical skills to changes in teachers’ pedagogical practices relies on two assumptions. First, it assumes that teachers are thoughtful professionals with intentions
behind their decisions. Second, teachers’ actions are a result of their thoughts and intentions (Shavelson & Stern, 1981). Based upon these assumptions, a simplified version of Shavelson and Stern’s (1981) theoretical model, shown in Figure 2 below, helps explain why these connections hold.

**Figure 2:** Theoretical framework connecting changes in pedagogical knowledge to changes in pedagogical decisions

![Theoretical Framework Diagram]

**Theory Underlying Sense of Self as a Math Teacher/Math Efficacy**

Informing teachers of the various pedagogical strategies, such as cooperative learning and classroom management, and of the reasons why they work should change teachers’ perspectives of the nature of instructional tasks. For example, most teachers teach math in the manner in which they were taught: they place students in orderly rows, give them steps/examples on how to solve a problem, and then give them opportunities for individual practice (Putnam & Borko, 2000). Introducing these teachers to cooperative learning by providing theories and evidence that knowledge and the
conception of ideas are results of group interactions and that learning is a social process (Putman & Borko, 2000) has now given the teachers a new idea about the nature of learning. Additionally, by giving the teachers practical grouping strategies (e.g.: heterogeneous, homogenous), it has also provided means for broadening the teachers’ ideas of the nature of instructional tasks. This, in turn, may now affect a teachers’ judgment as to the manner in which they have traditionally been teaching; they may now hold the perspective that students may be more inclined to learn through cooperative groups, thus change the way students are seated in the classroom.

Pedagogical knowledge also affects how teachers attribute the causes of student behavior. For instance, teachers may attribute poor student behavior and interest solely to the students. However, an effective professional development initiative may inform the teachers that “engagement rates depend on the teacher’s ability to organize the classroom as an efficient learning environment where activities run smoothly, transitions are brief and orderly, and little time is spent getting organized or dealing with misconduct” (Brophy, 1986, p 1070). This may convince teachers that effective classroom management strategies could curtail poor student behavior and engagement and thus change their pedagogical decisions and practices.

Beliefs held by teachers influence their judgments, perceptions, and attitudes, which in turn affect their behavior and classroom practices (Aguirre & Speer, 2000; Artzr & Armour-Thomas, 2002; Garmon, 2004; Pajares, 1992). That is, beliefs function as mediators between teachers’ knowledge (pedagogical and content) and their goals and practices (Artzr & Armour-Thomas, 2002). Thus, a change in goals and practices could
be due to a change in beliefs. Or, conversely, a change in beliefs leads to change in teachers’ goals and practices (Aguirre & Speer, 2000).

For example, a teacher with low sense of self as a math teacher may feel that she does not have a good understanding of the math that she is teaching or that she does not teach mathematics well. Her goals, then, would be to stay as close as possible to the curriculum as it is presented in the book and not to allow too much discussion so that her perceived inadequacies do not become explicit to her students. Her classroom practices reflect this by using a “teacher as expert” model, in which she demonstrates the steps in solving the problems, and then assigns similar and non-challenging problems to her students. If professional development were to increase this teacher’s mathematical self-efficacy so that she now feels that she is better able to handle the mathematical content, perhaps her goals could be switched from hiding her perceived inadequacies to having her students understand the content. Her practices may reflect her changed self-perceptions by allowing students to discuss possible solutions and by assigning more challenging problems.

The above explication of the mechanisms by which a teacher’s sense of self can be induced to change provides important support for assuming that professional development focused on changing this belief could positively impact student achievement. In an study of pre-service teachers’ beliefs in terms of multicultural awareness and sensitivity, Garmon (2004) identified two general categories of factors that influence positive changes in teacher beliefs: dispositional and experiential. Dispositional characteristics (such as being open to new ideas and/or self-reflective) are difficult or impossible to change through interventions. On the other hand, experiential factors (such
as education, successes as a math teacher, or other personal experiences) are more malleable and could be changed through appropriate professional development. Thus, the theoretical framework shown in Figure 3 for changing teacher behavior through professional development arises. Although no research was found that expanded Garmon’s framework from multicultural/equity beliefs to general beliefs, such as math efficacy, this framework provides a way of thinking about other beliefs that teachers might hold.

**Figure 3:** Theoretical framework connecting changes in beliefs to changes in teacher practices

```
Theory Underlying Shared Educational Values

The model in Figure 3 illustrates that similar experiences (such as through professional development) could induce teachers to share similar beliefs, thus similar goals and practices. If the experiences influence individual teachers’ beliefs in positive ways, then the students of these individual teachers are more likely to be exposed to better practices and, consequently, to improve their achievement.
However, the model in figure 3 does not explain why creating a shared vision among the teachers in a school would lead to changes in student achievement. Pearce and Ensley (2004) argued that a shared vision should lead to greater gains in student achievement than simply a single teacher changing his or her beliefs:

At the individual level, cognition and understanding of goals have been shown to positively affect confidence [...]. Since shared vision entails greater understanding and concrete agreement, by all [teachers], of the task at hand it seems likely that shared vision should affect the level of efficacy the [teachers] feel toward the potential of their innovation (pp. 262-263).

Pearce and Ensley also discussed how a shared vision should affect motivation and teamwork. When shared visions are created, individuals feel that they are being held responsible for their contributions, thereby increasing their extrinsic motivation to succeed. In other words, if the shared vision among teachers in a school is to have a certain percentage of the students reach proficiency, then each individual teacher will feel responsible for helping the “team” or school by completing his or her role in the successful completion of that goal. Moreover, if teachers fully understand their shared vision, then it seems reasonable that they would also feel more empathetic to their colleagues’ situations. Also, they would understand how their tasks are inter-related in the completion of the goal. Thus, teachers may be more helpful towards one-another, which would facilitate sharing content and pedagogical knowledge. This, in turn, would lead to increased student achievement through the mechanisms described earlier.

*Theory Underlying Equity*

Much of the literature on equity is based on the theoretical framework that differences in race and SES can cause cultural discontinuity between teachers and
students. Cultural discontinuity can be described as differences between a child’s home environment (including values and cultural norms) and his/her school environment (Cholewa & West-Olatunji, 2008; Ogbu, 1982). Most elementary-school teachers are White with middle-class backgrounds; what is more, many minority teachers have been educated and assimilated into the dominant White, middle-class culture. As a result both White and minority teachers often hold a similar set of norms that differ from the shared norms of their minority and low SES students.

When teachers are faced with cultural discontinuity between themselves and their students, their reaction often creates an obstacle for student learning (Graybill, 1997). In short, cultural discontinuity affects student achievement. For example, students enter into the school system armed with the learning and behavioral characteristics of their home cultures, teachers often interpret these aspects of their students’ culture as misbehavior in the classroom which causes the teacher to have lowered expectations and can even lead to inappropriately referring the student to special education services (Hoover, 2009).

Although professional development cannot reduce the cultural discontinuity between teachers and students of low-SES backgrounds, English language learners, or minorities, it can provide information and opportunities for teachers to learn about these students’ cultural norms. Many educational researchers have suggested that teachers must have an understanding of the cultural orientations of their students and their students’ families in order to increase student achievement (Hoover, 2009; Kea, Campbell-Whatley, & Bratton, 2008; Ladson-Billings, 2006). Kea, Campbell-Whatley, and Bratton (2008) have argued that teachers, not students, must adjust to the culture in the classroom by becoming more self-reflective and informed, instead of simply requiring students to
acculturate to the school norms. With high-quality professional development that emphasizes the importance of recognizing and respecting the values students bring with them to the classroom, teachers may be able to move away from a cultural deficit perspective, in which the values of minority and low SES students are viewed as inherently lacking, to a cultural difference perspective where contributions from all groups are equally respected (Graybill, 1997).

**Empirical Evidence in Support of the Categorization Scheme**

Very few teacher qualities have been definitively shown to correlate with student achievement. However, there is empirical support that increasing the seven teacher qualities chosen for inclusion in this study should increase students’ math achievement.

*Subject Matter Knowledge (SMK)*

The trend to measure teachers’ content knowledge through performance on math assessments rather than through level of education, certification, or general knowledge exams began in the 1990s (Hill, Sleep, Lewis, & Ball, 2007). Many studies administered the same tests as would be given to the teachers’ math students (e.g.: Harbison and Hanushek, 1992; Mullens, Murnane, & Willett, 1996). Thus, these studies looked at the extent to which teachers were able to perform the math problems they were teaching their students and the correlation between their performance on these tests and their students’ performance. Overall, these studies provide rather consistent evidence that higher teacher scores on math tests correlate with higher student scores (Hill et al., 2007).
Pedagogical Content Knowledge (PCK)

Shulman (1986a, 1986b) coined this term to represent the intersection between content and pedagogy. It situates the content that the teachers should know within teaching practices. In other words, there is some sort of mathematical knowledge specific to how to teach math that is different than how to do math. Shulman specifies that teachers who possess high PCK: a) know both how and why the content is being taught, b) are able to transform their content knowledge into effective pedagogy, c) represent content in various forms, d) understand level of difficulty of the content for students, and e) identify and correct student misconceptions. Carpenter, Fennema, Peterson, and Carey (1988) conducted one of the first studies to make explicit use of Shulman’s concept of PCK. First grade teachers who were able to identify which problems their students would be able to successfully complete had students with higher math achievement scores. More recently, Hill, Rowan, and Ball (2005) also found correlations between first and third grade teachers’ mathematical PCK and their students’ math achievement. Other studies (Ball & Bass, 2002; Ball & Bass, 2003; Hill, Schilling, & Ball, 2004) have identified teacher practices that make use of Shulman’s idea of PCK. Although these last studies have not made direct connections to student achievement, the authors argue that teachers whose practices are consistent with Shulman’s PCK are also highly effective teachers.

One aspect of PCK is knowing how students learn. There are two different approaches to learning how students learn: eliciting students to think out loud in order for the teacher to understand their thought processes or understanding the theories behind how students learn. Perhaps one of the best examples of having teachers learn students’ thought processes while solving math problems is Cognitively Guided Instruction (CGI),
developed by Carpenter, Fennema, Peterson & Franke (Promising Practices Network, 2007). The research on CGI has shown positive and significant effects on elementary students’ math achievement (Carpenter, Fennema, Franke, 1996; Carpenter, Fennema, Franke, Levi, & Empson, 2000). In a slightly different direction, a correlation has been found between teachers who have higher knowledge in learning theories, such as constructivism, and student achievement (Staub & Stern, 2002; Abbott & Fouts, 2003). This would imply that learning how students learn from a learning-theory perspective, as opposed to CGI’s eliciting student’s thinking point of view, may also have an effect on student achievement.

**Equity**

In 2000, Skrla, Scheurich, and Johnson published a report on several school districts that met with district-wide success in increasing the academic performance of minority and low-SES students. The concepts implemented by these districts were termed “equity-driven” and “achievement-focused;” these concepts were integrated into all levels in the district, from the school boards to the classrooms. The study found that implementing equity-focused district-wide programs increased student achievement. The programs also resulted in a change to an “equity-driven” pedagogy at the classroom level, which resulted in this increased achievement. Equity-focused professional development holds promise as a possible moderator of professional development’s effect on student achievement.
Creating a Shared Vision

Results consistently suggest that “the implementation of a clear mission, shared vision, and school-wide goals promote increased student achievement” (Stolp, 1994, p 2). Whether the shared vision comes as a result of teacher professional groups (Louis & Marks, 1998; Morrissey, 2000) or through an administrator’s leadership (Weiss & Cambone, 1994), it has been show to produce effective school reforms.

Increasing Sense of Self as a Math Teacher

A teacher’s strong (or weak) sense of self as a math teacher has been shown to have long-lasting impacts on their students. Studies have traced the evolution of mathematics anxiety among high school and college students to their elementary school classroom experiences. This anxiety may have been exacerbated by their teachers since mathematically anxious teachers have been shown to pass their anxieties on to their students (Buhlman & Young, 1982), thus undermining their roles as effective or inspiring teachers of math. Moreover, mathematically anxious teachers tend to rely more on traditional teaching methods and basic skills rather than concept development (Brush, 1981). These methods run contrary to the recommendations of the National Council of Teachers of Mathematics (NCTM) (1989) that has stressed the importance of problem solving methods and collaborative learning environments. As such, increasing teachers’ sense of self as math teachers should decrease these problems related to mathematical anxiety and help teachers align their teaching to that standards supported by the NCTM. Goddard, Hoy, and Hoy (2000) found that teachers with higher self-efficacy (a concept closely related to sense of self) have students with higher achievement.
Cooperative Learning

Cooperative learning has been an ongoing hot topic in education and has received much attention by researchers, especially in the 1980s and 1990s (e.g.: Sharon, 1980; Slavin, 1980; Slavin, 1983; Johnson & Johnson, 1999; Brown & Palincsar, 2000). In his 1991 research synthesis, Slavin determined that when groups have a set of goals and hold individual students accountable, cooperative learning resulted in increased student achievement. The positive effects of cooperative learning on students’ math achievement were again confirmed in Slavin and Lake’s (2008) research synthesis discussed above.

Classroom Management

Good classroom management creates an environment that supports teaching and learning (McGarity & Butts, 1984; Brophy, 1988; McGinnis, Frederick, & Edwards, 1995). Often, the phrase “classroom management” is confounded with “classroom discipline”. However, classroom management for the above studies has been defined in much broader terms. For example, McGarity and Butts (1984) included how teachers identify students who need individual help, maintain active participation, and use time effectively in addition to how teachers deal with behavioral problems. “Classroom management” viewed through this wider lens has been shown to have significant positive effects on student achievement.

The above sections have demonstrated that there exists theoretical and empirical evidence supporting an investigation into these foci of professional development as
predictors of student gains in mathematics. Instead of examining the effects of these foci individually and then synthesizing the results, this study seeks to quantitatively compare the effects. In order to do so, meta-analytic techniques were employed and are described in detail below.
The methods of a meta-analysis mirror that of any quantitative research study. First, a population is identified; for a meta-analysis, this consists of searching the literature and determining the inclusion and exclusion criteria of the studies. Next, data are collected by a specific coding process by which important information is gathered from each study. Once the relevant data have been collected, they are analyzed and results are found. The following sections discuss each of these steps in detail.

Searching the Literature

The literature search took place between August and November of 2009. The following flowcharts (Figure 4) represent the process followed to identify relevant studies between 1989 and 2009 (i.e., the past two decades, since the inception of the mathematics reform/standards movement) in each of the databases related to education. The databases searched include: ERIC, Education Full Text, PsychINFO, Professional Development Collection, and Digital Dissertations Online. The searches were all done on the bases of Key Words. The key words used for this study were: Professional Development, Student Achievement, and Math*. Many of the databases contain thesauruses which advise the searcher about which terms are associated with each key word. These terms, which appear un-bolded in the flowcharts, were included in the search using the Boolean “or” to broaden search results. The Professional Development Collection and Digital Dissertations Online suggested the use of Teacher Education in place of Professional Development, so the search methods were adapted for these two databases. After the
main phase of each search, the titles and abstracts of the records were perused in order to ensure relevance. Records that included technology-specific reforms, comprehensive school reforms, and new curricular implementations were excluded due to the confounding nature of the variables. It would be very difficult, if not impossible, to parse out the effects of professional development separate from that of the technology, other aspects of the comprehensive reforms, or the new curriculum. Articles also were excluded if only pre-service teachers were included. Moreover, this meta-analysis focused only on the effects of professional development on the mathematic achievement of elementary students. Elementary school was defined as grades one through five. Grade six was included if it was part of an elementary school, but not if it was part of a middle school.

Following the electronic searches of the previous databases, an additional search was conducted to identify any reviews of literature or research syntheses concerning professional development. This search included known sources of educational research synthesis (e.g.: Review of Educational Research, Review of Research in Education, and handbooks of education) and Google Scholar. This process identified the three recent research syntheses looking at the effects of professional development on students’ math achievement discussed above. The references of these articles were perused and 104 citations were deemed possibly relevant based on their titles.
Figure 4: Flowchart of the Literature Search Process

**Database: ERIC**
(Date Range: 1989 – 2009)
(Age Group: Elementary)

**Professional Development**
- or Teacher Improvement or Faculty Development or Inservice Teacher Education
  - 2,890 records
  - AND

**Student Achievement**
- Or Academic Achievement
  - 234 records
  - AND

**Math**
- 51 records

Author Perusal of Titles and Abstracts
- 17 records

**Database: Education Full Text**
(Date Range: 1996 – 2009)

**Professional Development**
- or Teacher Education Inservice
  - 15,478 records
  - AND

**Student Achievement**
- 725 records
  - AND

**Math**
- 106 records

Author Perusal of Titles and Abstracts
- 27 records
**Database: PsychINFO**  
(Date Range: 1989 – 2009)  
(Age Group: School Age 6 - 12)

Professional Development  
or Inservice Teacher Education  

71 records  

AND  

**Student Achievement**†

or Academic Achievement  

3 records  

Author Perusal of Titles and Abstracts  

0 records

† Student Achievement was not recognized by the PsychINFO thesaurus, therefore this search only used Academic Achievement.

**Database: Professional Development Collection**‡‡
(Date Range: 1989 – 2009)

Teacher Education  

AND  

Academic Achievement  

104 records  

AND  

Math*  

12 records  

Author Perusal of Titles and Abstracts  

2 records

‡‡ Because of the nature of this database, the keyword search was adjusted to more efficiently search for records.

**Database: Digital Dissertations Online**‡+++  
(Date Range: 1989 – 2009)

Teacher Education  

AND  

Student Achievement  

AND  

Math*  

31 records  

Author Perusal of Titles and Abstracts  

1 record

‡+++ Because of the limited search capabilities of this database, the Keyword search needed to take place simultaneously and synonyms were excluded.
This search resulted in 28 studies which met the specific inclusion criteria. They fell into five separate study designs: 3 regression studies (either multiple regression or hierarchical linear modeling), 2 studies which report the percent increase of proficient students, 3 studies with control groups which report only the post-test scores, 14 studies with control groups which report both the pre-test and post-test scores, and 6 studies that report pre-test and post-test scores but have no control group (i.e., they only report gain scores). An effort was made to obtain more studies by contacting researchers who had been cited for studying professional development interventions. Two researchers replied and provided data from their studies.

For the quantitative analysis portion of this meta-analysis, only the 17 studies (the 14 pre- post-test and control studies and the 3 post-test with control studies) which had control groups were included. This was necessary for two reasons. First, effect sizes for these different types of studies/analysis methods are each calculated in a different manner (Hedges, 2009). Doing this would lead to effect sizes that are cannot be compared (Lipsey & Wilson, 2001). Also, the quality of the results between these types of studies is not equivalent. Specifically, gain score studies may conclude that there was an effect of professional development based on the differences between pre-test and post-test scores. However, without a proper control group it is impossible to tell whether the growth was specifically due to the professional development intervention or due to other unknown factors. Thus, using only those studies with a control/comparison group made it possible to assume causation between the professional development and differences in student achievement.
Gathering Information from the Studies

I developed a coding sheet to serve as the template for collecting and coding pertinent information about each study. Broken into subsections, the coding sheet helped identify information concerning article characteristics (e.g.: relevance, quality of study), effect size data, professional development characteristics (e.g.: focus, duration), and outcome measures characteristics, as well teacher and student demographics. The coding sheet and effect size data collection table are included in Appendix A and Appendix B, respectively.

The operationalization process for determining the focus of the professional development was based on the description of the professional development provided by each article’s authors. As a result, if the authors’ description of the professional development was inaccurate, then the validity of the results of this meta-analysis is threatened. One solution to this validity threat would be to gather the curricular/training materials used to implement the professional development and directly assess the focus from these materials. Although this may not necessarily be a perfect representation of the enacted professional development, it would at least be a representation of the intended purpose. The gathering of the curricular materials was beyond the scope of this study; consequently, the validity of these results lies partially in the hands of the authors of the included studies. It is possible that at some point in the future, these curricular materials could be gathered and the reliability of this coding process could be assessed. Perhaps, some combination of the author’s description of the enacted intervention, which is being used in this study, and the intended intervention would constitute the best measure of the focus of the professional development interventions.
Coding was conducted by both the author and another doctoral student studying math education. We met to discuss the coding sheet and to clarify any points of confusion. The source of most confusion was the classification of the seven different professional development foci. This led to the foci being specifically operationalized as follows:

*Subject Matter Knowledge:* If the focus of the professional development was to increase teachers’ knowledge of mathematical concepts, then it was coded for having an SMK focus. The presentation of math concepts would have to occur separate from instruction as to how to teach these concepts. One way of determining if increasing content knowledge was a focus is the inclusion of an assessment of teachers’ mathematical knowledge as part of the study.

*Pedagogical Content Knowledge:* Professional development that focused on how mathematical concepts should be taught was coded as having a PCK focus. PCK often, but not always, includes teaching multiple ways of representing the same concept and how to use these multiple representations to help students learn. To help differentiate, teachers learning how to compute fraction problems would fall under the Subject Matter Knowledge focus while teachers learning how to teach fraction problems would fall under the PCK focus. An intervention that first focused on how to solve fractions and then applies this knowledge to how to teach fractions was coded as having both a Subject Matter Knowledge and a PCK focus.

A focus on learning how students learn was also coded as a focus on PCK. A focus on learning discusses what students are thinking as they learn a concept. This can
come in several forms. One form can have teachers learning about various learning theories (e.g.: constructivism, behaviorism, etc.). Another form can outline specifically what students may be thinking as they solve math problems, including misconceptions they might be applying (e.g.: Cognitively Guided Instruction).

Equity: Professional development was said to have an equity focus if it was created or designed to target a specific under-represented population. These populations can include special education students, minorities, students from low socio-economic backgrounds, bilingual students, or students labeled “at-risk”. The equity focus will rarely occur on its own and will mostly likely be paired with another focus. This focus may or may not be independent of content. For example, teachers learning about their students’ culture and teachers learning how to teach math to special education students can both be coded for equity.

Creating a Shared Vision: If teachers were asked to work towards creating a unified view of what to teach, how to teach, their beliefs about learning, etc., then the focus of the professional development was coded as “creating a shared vision.” For example, a professional development intervention which had teachers work together to determine where to place curricular emphasis, then creating a shared vision was a focus of this intervention.

Increasing Sense of Self as a Math Teacher (Math Efficacy): In order for Math Efficacy to be the focus of professional development, the study must have explicitly
stated this in their description; it cannot be implied or a post-hoc side effect. Phrasing for this included “increasing teachers’ sense of self”, “increasing teachers’ self-efficacy”, “increasing teachers’ confidence”, etc.

**Cooperative Learning Strategies**: Professional development which helped teachers determine how to activate their students to help each other would fell under the Cooperative Learning focus. This included both grouping strategies (e.g.: heterogeneous vs. homogeneous) or peer-tutoring strategies. Cooperative Learning is independent of content. That is, informing teachers how to teach/organize their lesson with students in groups counted only for Cooperative Learning and not for PCK because the PCK focus required a specific content component (i.e., how to teach Math, not how to teach groups).

**Classroom Management**: Professional development with classroom management as the focus helped teachers avoid discipline problems and create a learning environment with supportive and attentive students. Like Cooperative Learning, this focus is entirely extraneous to any content.

After coming to common agreement on how to operationalize the terms on the coding sheet, both the author and doctoral student coded each study. A comparison of the coding database determined that only a few differences existed in the information that was retrieved. In almost all cases, the differences occurred when determining the focus of the professional development being studied. The inter-rater agreement for variables such as Article Characteristics and Effect Size Data was often perfect or close to perfect.
However, the initial inter-rater agreement for the focus variables was approximately 85%. This was probably due to the need to impugn the professional development’s content focus from textual descriptions that ranged from short to very long; in contrast, authors’ names and student achievement data can be read directly from the article. Differences were resolved through further discussion until an agreement was reached. The following tables highlight some of the important information gathered from the studies. QED refers to a quasi-experimental design while RCT refers to a randomized control trial.
Table 1: List of identified studies and key study characteristics

<table>
<thead>
<tr>
<th>Study</th>
<th>Source</th>
<th>PD_Name</th>
<th>Focus</th>
<th>Study Design</th>
<th>Grades</th>
<th>Treatment Student N (All Students N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Burkhouse, Loftus, Sadowski, &amp; Kathy (2003)</td>
<td>Report</td>
<td>Thinking Mathematics (TM)</td>
<td>PCK</td>
<td>Survey</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Cardelle-Elawar (1990)</td>
<td>Journal</td>
<td>None</td>
<td>PCK &amp; Equity</td>
<td>QED</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>Cardelle-Elawar (1995)</td>
<td>Journal</td>
<td>Metacognitive Instruction</td>
<td>PCK</td>
<td>RCT</td>
<td>3,4,5</td>
</tr>
<tr>
<td>4</td>
<td>Carpenter, Fennema, Peterson, Chiang, &amp; Loef (1989)</td>
<td>Journal</td>
<td>Cognitively Guided Instruction</td>
<td>PCK</td>
<td>QED</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>Freiberg, Prokosch, Treister, &amp; Stein (1990)</td>
<td>Journal</td>
<td>Consistency Management</td>
<td>Shared Vision &amp; Classroom Mgmt</td>
<td>QED</td>
<td>1,2,3,4,5.</td>
</tr>
<tr>
<td>8</td>
<td>Fuchs, Fuchs, Karns, Hamlett, &amp; Katzaroff (1999)</td>
<td>Journal</td>
<td>Performance Assessment</td>
<td>PCK</td>
<td>QED</td>
<td>2,3,4</td>
</tr>
<tr>
<td></td>
<td>Authors</td>
<td>Year</td>
<td>Study Type</td>
<td>Title</td>
<td>SMK &amp; PCK</td>
<td>QED</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------</td>
<td>------</td>
<td>------------</td>
<td>----------------------------------------------------------------------</td>
<td>-----------</td>
<td>-----</td>
</tr>
<tr>
<td>9</td>
<td>Jagielski (1991)</td>
<td></td>
<td>Dissertation</td>
<td>Mathematical Curriculum Improvement Project (MCIP)</td>
<td>SMK &amp; PCK</td>
<td>QED</td>
</tr>
<tr>
<td>11</td>
<td>Martin (2006)</td>
<td></td>
<td>Dissertation</td>
<td>None</td>
<td>SMK &amp; PCK</td>
<td>QED</td>
</tr>
<tr>
<td>12</td>
<td>Mason &amp; Good (1993)</td>
<td></td>
<td>Journal</td>
<td>None</td>
<td>Cooperation Learning</td>
<td>QED</td>
</tr>
<tr>
<td>14</td>
<td>Ross, Bruce, &amp; Hogaboam-Gray (2006)</td>
<td></td>
<td>Journal</td>
<td>Reform Communication</td>
<td>PCK</td>
<td>RCT</td>
</tr>
<tr>
<td>15</td>
<td>Sloan (1993)</td>
<td></td>
<td>Dissertation</td>
<td>Direct Instruction Model</td>
<td>PCK</td>
<td>QED</td>
</tr>
<tr>
<td>16</td>
<td>Villasenor &amp; Kepner (1993)</td>
<td></td>
<td>Journal</td>
<td>Cognitively Guided Instruction</td>
<td>PCK</td>
<td>QED</td>
</tr>
</tbody>
</table>

*Sample size refers to number of classes instead of number of students.
NOTE: QED stands for Quasi-Experimental Design. RCT stands for Randomized Control Trial.
Table 2: Measures of student math achievement

<table>
<thead>
<tr>
<th>Study</th>
<th>Measure (Subtests)</th>
<th>Reliability</th>
<th>Pre-test Mean (SD) Treatment Group</th>
<th>Pre-test Mean (SD) Control Group</th>
<th>Post-test Mean (SD) Treatment Group</th>
<th>Post-test Mean (SD) Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burkhouse, Loftus, Sadowski, &amp; Kathy (2003)</td>
<td>PSSA</td>
<td>&lt; .90</td>
<td>N/A</td>
<td>N/A</td>
<td>1307(164.16)</td>
<td>1216(178.67)</td>
</tr>
<tr>
<td>Cardelle-Elawar (1990)</td>
<td>Research Designed (Parellel Forms: Males)</td>
<td>0.76 &amp; 0.81</td>
<td>5.30(1.87)</td>
<td>5.25(1.56)</td>
<td>12.00(2.20)</td>
<td>6.20(1.17)</td>
</tr>
<tr>
<td>Cardelle-Elawar (1990)</td>
<td>Research Designed (Parellel Forms: Females)</td>
<td>0.76 &amp; 0.81</td>
<td>5.45(1.47)</td>
<td>4.75(1.74)</td>
<td>10.60(1.98)</td>
<td>5.90(1.91)</td>
</tr>
<tr>
<td>Cardelle-Elawar (1995)</td>
<td>Research Designed (3rd Grade)</td>
<td>Not Available</td>
<td>1.56(.50)</td>
<td>1.85(2.00)</td>
<td>15.00(1.30)</td>
<td>8.32(1.00)</td>
</tr>
<tr>
<td>Cardelle-Elawar (1995)</td>
<td>Research Designed (4th Grade)</td>
<td>Not Available</td>
<td>1.80(.86)</td>
<td>2.35(1.96)</td>
<td>14.60(1.95)</td>
<td>9.50(1.90)</td>
</tr>
<tr>
<td>Cardelle-Elawar (1995)</td>
<td>Research Designed (5th Grade)</td>
<td>Not Available</td>
<td>2.60(1.50)</td>
<td>3.50(1.70)</td>
<td>16.00(3.20)</td>
<td>11.10(2.60)</td>
</tr>
<tr>
<td>Carpenter, Fennema, Peterson, Chiang, &amp; Loef (1989)</td>
<td>ITBS (Level 7) (Computation)</td>
<td>0.89</td>
<td>N/A</td>
<td>N/A</td>
<td>20.95(2.08)</td>
<td>20.05(1.81)</td>
</tr>
<tr>
<td>Carpenter, Fennema, Peterson, Chiang, &amp; Loef (1989)</td>
<td>ITBS (Level 7) (Problem Solving)</td>
<td>0.9</td>
<td>N/A</td>
<td>N/A</td>
<td>17.28(1.83)</td>
<td>16.42(1.89)</td>
</tr>
<tr>
<td>Freiberg, Connell, &amp; Lorentz (2001)</td>
<td>TAAS</td>
<td>Not Available</td>
<td>40.58(16.76)</td>
<td>44.23(16.79)</td>
<td>49.08(19.63)</td>
<td>45.62(18.56)</td>
</tr>
<tr>
<td>Freiberg, Huizinec, &amp; Templeton (2009)</td>
<td>TAAS</td>
<td>Not Available</td>
<td>75.44(10.49)</td>
<td>75.74(19.67)</td>
<td>81.80(8.41)</td>
<td>78.03(9.60)</td>
</tr>
<tr>
<td>Fuchs, Fuchs, Karns, Hamlett, &amp; Katzaroff (1999)</td>
<td>Research Designed (Conceptual Understanding)</td>
<td>0.56</td>
<td>1.22(.54)</td>
<td>1.20(.56)</td>
<td>2.70(1.27)</td>
<td>1.35(0.72)</td>
</tr>
<tr>
<td>Study</td>
<td>Instrument/Domain</td>
<td>Measure 1</td>
<td>Measure 2</td>
<td>Measure 3</td>
<td>Measure 4</td>
<td>Measure 5</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>------------------------------------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>Fuchs, Fuchs, Karns, Hamlett, &amp; Katzaroff (1999)</td>
<td>Research Designed (Computation)</td>
<td>0.66</td>
<td>.61(.65)</td>
<td>.49(.65)</td>
<td>2.12(1.36)</td>
<td>.86(.95)</td>
</tr>
<tr>
<td></td>
<td>Research Designed (Problem Solving)</td>
<td>0.59</td>
<td>.73(.71)</td>
<td>.65(.75)</td>
<td>2.41(1.29)</td>
<td>1.04(.93)</td>
</tr>
<tr>
<td></td>
<td>Research Designed (Communication)</td>
<td>0.46</td>
<td>.91(.61)</td>
<td>.93(.70)</td>
<td>2.76(1.25)</td>
<td>1.31(.91)</td>
</tr>
<tr>
<td>Karper &amp; Melnick (1993)</td>
<td>Derry Township Standardized Test (3rd Grade)</td>
<td>Not Available</td>
<td>N/A</td>
<td>N/A</td>
<td>64.9(100)</td>
<td>73.8(100)</td>
</tr>
<tr>
<td></td>
<td>Derry Township Standardized Test (4th Grade)</td>
<td>Not Available</td>
<td>63.30(100)</td>
<td>63.30(100)</td>
<td>73.00(100)</td>
<td>73.00(100)</td>
</tr>
<tr>
<td></td>
<td>Derry Township Standardized Test (5th Grade)</td>
<td>Not Available</td>
<td>65.50(100)</td>
<td>65.50(100)</td>
<td>70.30(100)</td>
<td>66.30(100)</td>
</tr>
<tr>
<td></td>
<td>Research Designed (4th Grade; Year 2)</td>
<td>Not Available</td>
<td>15.66</td>
<td>19.10</td>
<td>29.33(8.40)</td>
<td>25.68(5.90)</td>
</tr>
<tr>
<td></td>
<td>Research Designed (4th Grade; Year 3)</td>
<td>Not Available</td>
<td>17.65</td>
<td>16.84</td>
<td>30.05(7.95)</td>
<td>25.68(8.56)</td>
</tr>
<tr>
<td></td>
<td>Research Designed (5th Grade; Year 1)</td>
<td>Not Available</td>
<td>8.63</td>
<td>7.65</td>
<td>21.24(8.67)</td>
<td>18.21(7.84)</td>
</tr>
<tr>
<td></td>
<td>Research Designed (5th Grade; Year 2)</td>
<td>Not Available</td>
<td>16.65</td>
<td>19.17</td>
<td>26.54(7.27)</td>
<td>27.18(5.02)</td>
</tr>
<tr>
<td></td>
<td>Research Designed (5th Grade; Year 3)</td>
<td>Not Available</td>
<td>16.99</td>
<td>18.64</td>
<td>27.13(7.49)</td>
<td>25.39(7.24)</td>
</tr>
<tr>
<td>Mason &amp; Good (1993)</td>
<td>Research Designed (Computation)</td>
<td>0.89 (overall avg.)</td>
<td>N/A</td>
<td>N/A</td>
<td>.18(.97)</td>
<td>-.14(1.00)</td>
</tr>
<tr>
<td>Mason &amp; Good (1993)</td>
<td>Research Designed (Conceptual Understanding)</td>
<td>0.89 (overall avg.)</td>
<td>N/A</td>
<td>N/A</td>
<td>.11(.97)</td>
<td>-.04(1.00)</td>
</tr>
<tr>
<td>Study</td>
<td>Design Type</td>
<td>Metric</td>
<td>Overall Avg</td>
<td>SD 1</td>
<td>SD 2</td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>-------------------------</td>
<td>-----------------</td>
<td>-------------</td>
<td>------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>Mason &amp; Good (1993)</td>
<td>Research Designed</td>
<td>Problem Solving</td>
<td>0.89</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Mason &amp; Good (1993)</td>
<td>Research Designed</td>
<td>Estimation</td>
<td>0.89</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Pace (2008)</td>
<td>CSAP</td>
<td>3rd Grade</td>
<td>Not Available</td>
<td>N/A</td>
<td>N/A</td>
<td>500.35(80.38)</td>
</tr>
<tr>
<td>Pace (2008)</td>
<td>CSAP</td>
<td>4th Grade</td>
<td>Not Available</td>
<td>N/A</td>
<td>N/A</td>
<td>534.28(72.92)</td>
</tr>
<tr>
<td>Pace (2008)</td>
<td>CSAP</td>
<td>5th Grade</td>
<td>Not Available</td>
<td>N/A</td>
<td>N/A</td>
<td>543.56(73.53)</td>
</tr>
<tr>
<td>Sloan (2006)*</td>
<td>EQAO</td>
<td>Not Available</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Villasenor &amp; Kepner (1993)</td>
<td>Researcher Designed</td>
<td>Word Problems</td>
<td>Not Available</td>
<td>N/A</td>
<td>N/A</td>
<td>5.44(.61)</td>
</tr>
<tr>
<td>Villasenor &amp; Kepner (1993)</td>
<td>Researcher Designed</td>
<td>Number Facts</td>
<td>Not Available</td>
<td>N/A</td>
<td>N/A</td>
<td>4.86(.56)</td>
</tr>
<tr>
<td>Villasenor &amp; Kepner (1993)</td>
<td>Researcher Designed</td>
<td>Arithmetic</td>
<td>Not Available</td>
<td>N/A</td>
<td>N/A</td>
<td>2.26(0.56)</td>
</tr>
<tr>
<td>Walsh-Cavazos (1994)</td>
<td>Research Designed</td>
<td></td>
<td>0.74</td>
<td>57.00(16.36)</td>
<td>51.92(16.31)</td>
<td>66.38(16.73)</td>
</tr>
</tbody>
</table>

* Reported t-test statistics
Data Analysis

Calculating Effect Sizes

To answer the research questions concerning the effects of the foci of professional development on elementary students’ math achievement, meta-analytic techniques were employed. As opposed to a literature review or narrative research synthesis, “a meta-analysis is systematic and organized, and typically uses statistical analysis to try and make sense of the series of results” (Hedges, 2009). In other words, instead of qualitatively describing the literature, a meta-analysis quantitatively synthesizes the results of all relevant previous studies in a systematic manner. The strengths of a meta-analysis lie with its ability to synthesize the results of many studies, which makes the results more powerful than those of a single study, and with the ability to conduct moderation analysis (Lipsey & Wilson, 2001). To do this, the results from the studies are transformed into a standardized metric known as the effect-size.

For the sake of clarity, the following will be used as general notations throughout the data analysis:

- A subscript $i$ refers to the $i$th effect size, such that $i = 1, 2, \ldots, k$ and $k$ is the number of effect sizes.
- A subscript $T$ refers to the Treatment Group in the relevant study.
- A subscript $C$ refers to the Control Group in the relevant study.

This meta-analysis employed the standardized mean difference (a.k.a.: Cohen’s $d$-index) measure of effect size (Rosenthal, 1994).

$$g_i = \frac{\bar{x}_T - \bar{x}_C}{SD_{pooled\ i}}$$  \hspace{1cm} (1)

where
\( \bar{X}_T_i \) and \( \bar{X}_C_i \) = the two group means for the Treatment and Control groups, respectively. For cases in which pre- and post-test means are provided, \( \bar{X}_T_i \) and \( \bar{X}_C_i \) refer to the gain scores for the treatment and control groups.

\[ SD_{pooled_i} = \text{the estimated common standard deviation of the two groups given by:} \]

\[ SD_{pooled_i} = \sqrt{\frac{(n_T_i-1)SD_T^2 + (n_C_i-1)SD_C^2}{n_T_i + n_C_i - 2}} \]  

(2)

where

\( SD_T_i \) and \( SD_C_i \) = the standard deviation of the post-test scores for the Treatment and Control groups. However, if the pre- and post-test standard deviations vary greatly, \( SD_{pooled_i} \) should be calculated using both pre- and post-test standard deviations (Morris, 2008):

\[ SD_{pooled_i} = \sqrt{\frac{(n_T_i-1)SD_{preT_i}^2 + (n_C_i-1)SD_{preC_i}^2 + (n_T_i-1)SD_{postT_i}^2 + (n_C_i-1)SD_{postC_i}^2}{2(n_T_i + n_C_i - 2)}} \]  

(3)

where,

\( SD_{preT_i} \) and \( SD_{preC_i} \) = the standard deviation of the pre-tests of the Treatment and Control groups, and

\( SD_{postT_i} \) and \( SD_{postC_i} \) = the standard deviation of the post-tests of the Treatment and Control groups.

Often, studies did not provide the means, standard deviations, and sample sizes. If they provided the results from a \( t \)-test comparison and the corresponding degrees of freedom, the \( d \)-effect size was calculated using (Rosenthal, 1994):

\[ g_i = \frac{2t_i}{\sqrt{n_T_i + n_C_i - 2}} \]  

(4)
In other cases, the data provided by studies was aggregated by class or schools. Because individual students within a class (or school) tend to resemble each other more than those between schools, the aggregated means result in much higher effect sizes compared to effect sizes calculated for individual students (Hedges, 2009). In order to be able to compare these aggregated means with the non-aggregated means, an intra-class correlation (ICC) was applied to correct for unit differences when calculating the aggregated data’s effect sizes. Essentially, an ICC is the degree to which individuals in a group resemble each other. Following the guidelines specified by the What Works Clearinghouse (2008), this study used ICC = .20 as a default unless there were details provided in the article specifying another ICC. Thus, the effect size adjusted for the ICC correlation is given by:

$$g_{\text{i-adjusted}} = g_i \sqrt{ICC}$$  \hspace{1cm} (5)

Lastly, because effect sizes can be biased by the sample size (i.e., studies with larger samples will have more weight than studies with fewer samples), the original effect size was unbiased via the following formula (Hedges, 2005).

$$d_i = g_i \left[1 - \frac{3}{4(n_i-2)-1}\right]$$  \hspace{1cm} (6)

where

$$n_i = \text{the total sample size for that effect.}$$

From this point forward, all the $d$-effect sizes being presented and applied to future equations will have been unbiased.
Each effect size has an associated variance and 95% confidence interval (CI) given by:

\[ \nu_i = \frac{n_i}{n_T i \cdot n_C i} + \frac{d_i^2}{2n_i} \]  

and

\[ CI_{95\%} = d_i \pm 1.96\sqrt{\nu_i} \]  

\[ \text{Determining the Power of Detecting Effect Sizes} \]

Statistical power in a meta-analysis refers to the ability of the analysis to detect some level of effect and heterogeneity based on the number of effect sizes and the control and treatment sample sizes, as well as an assumed level of statistical significance, \( \alpha \). The assumed level of significance for this meta-analysis was set to \( \alpha = .05 \). The following description of a power analysis for overall effect sizes under a random effects model follows the work presented by Hedges and Pigott (2001).

The power of a two-tailed analysis that \( \Delta = \Delta_0 \) is given by:

\[ p = 1 - \Phi(c_{\alpha/2} - \lambda) + \Phi(-c_{\alpha/2} - \lambda) \]  

where,

\( \Phi(x) = \) the standard normal cumulative distribution function,

\( c_\alpha = \) the 100(1 - \( \alpha \)) percent point of the standard normal distribution function, and

\( \lambda = \) is the mean of the standard normal distribution function given by:

\[ \lambda = \frac{(\mu - \mu_0)}{\sqrt{\nu^*}} \]  

where,

\( \mu_0 \) is assumed to be zero, and
\[ \nu = \text{the population variance estimated by:} \]
\[ \nu = \frac{\nu_i + \hat{\tau}^2}{k} \]

(11)

and \( \nu_i \) is the estimated variance of the \( i \)th effect size, which is assumed equal for all effect sizes and is given as an average of the overall treatment and control sample sizes such that:
\[ \nu_i = \frac{\bar{n}_T + \bar{n}_C}{\bar{n}_T * \bar{n}_C} + \frac{d^2}{2(\bar{n}_T + \bar{n}_C)} \]

(12)

where,

\[ \bar{n}_T \text{ and } \bar{n}_C = \text{the average sample size for the treatment and control groups, and} \]

\[ d = \text{the chosen level of effect size that the power analysis is attempting to detect.} \]

Choosing \( d \) depends on what level of effect the analysis is attempting to detect. The three recent meta-analyses presented earlier were used as guidelines in choosing what effect sizes to expect, and thus, to attempt to detect. The effect sizes found in these meta-analyses range from 0.21 (Blank & de las Alas, 2009) to 0.38 (Scher & O’Reilly, 2009). Thus, the power of this meta-analysis to detect this range of effect sizes range from 0.81 to 1.00. The power of detecting any greater effect sizes will, understandably, be larger.

Finding the Overall Effect Size under a Random-effects Model

Once all effect sizes and their associated variances were calculated, it was necessary to identify whether or not there is a significant amount of variation between the effect sizes. From a descriptive standpoint, this determined whether or not the effect sizes differ. If they did not differ, then the effect of professional development on students’ math achievement was the same regardless of the quality or the focus (or any other
moderator) of the professional development; based on previous meta-analyses, this is probably not the case. If the effect sizes did differ, then it is possible to explore reasons to explain these differences, such as the focus of professional development. From a statistical standpoint, homogeneity testing was performed using the $Q$-statistic, which follows a $\chi^2$-distribution (Huedo-Medina, Marin-Martinez, Sanchez-Meca, & Botella, 2006) and is similar to running an ANOVA. $Q$ is given by:

$$Q = \sum_i w_i (d_i - \bar{d})^2$$

(13)

where

$$w_i = \frac{1}{v_i}$$

(14)

and

$$\bar{d} = \frac{\sum_i w_i d_i}{\sum_i w_i}$$

(15)

If the $Q$-statistic indicated that the effect sizes are not homogenous, a random-effects model was assumed for conducting the remaining analyses. In a random-effects model, the true effect sizes, or $\Delta_i$'s, are broken down into two components (Hedges & Vevea, 1998):

$$d_i = \Delta_i + e_i = \mu + \xi_i + e_i$$

(16)

where

$\mu =$ the mean effect size of the sampled population, and

$\xi_i =$ the error associated with estimating the $\mu$.

In simpler language: not all of the variance around the effect sizes can be attributed to sampling error, therefore, some of the variance is assumed to be due to differences in effect sizes. Thus, each of the effect sizes was assumed to be different and have its own error associated with it (i.e., a measure of uncertainty is being added to the model). The
variance associated with $\mu$ was computed by a formula represented by Hedges & Vevea (1998) as:

$$v_i^* = v_i + \tau^2$$  \hspace{1cm} (17)

where

$v_i$ = the variance associated with the $i$th effect size as given by equation (8), and

$\tau^2$ = the between-subject variance component.

There are two generally accepted methods to estimate the between variance component (Hedges & Vevea, 1998). The first, a function of the variance of the mean effect sizes and the mean of the variances, is given by:

$$\hat{\tau}_1 = \frac{\sum_{i=1}^{k} v_i}{k} + v_i$$  \hspace{1cm} (18)

The other, which is estimated based on the homogeneity test, is given by:

$$\hat{\tau}_2 = \begin{cases} 
\frac{Q-(k-1)}{c}, & Q \geq k - 1 \\
0, & Q < k - 1 
\end{cases}$$  \hspace{1cm} (19)

and

$$c = \sum_{i} w_i \frac{\sum_{i} w_i^2}{\sum_{i} w_i}$$  \hspace{1cm} (20)

where

$w_i$ = the inverse of the variance.

In order to be conservative with the estimation of the between-subject variance, both $\hat{\tau}_1$ and $\hat{\tau}_2$ are computed and the larger of the two was applied.

Once the between subject variance was computed, it was possible to compute the overall mean effect size for the random-effects model by using (Hedges & Vevea, 1998):

$$\bar{d}^* = \frac{\sum_{i} w_i^* d_i}{\sum_{i} w_i^*}$$  \hspace{1cm} (21)
where \( w_i^* \) is the weighing factor under the random-effects model and is given by:

\[
w_i^* = \frac{1}{v_i + \tau^2}
\]  

(22)

A Z-test can be used to test if the overall mean effect size under the random-effects model was significantly different from zero (i.e., test the null hypothesis \( H_0: \bar{d}^* = 0 \)). That is:

\[
Z = \frac{\bar{d}^*}{SE(\bar{d}^*)}
\]

(23)

where

\[
SE(\bar{d}^*) = \text{standard error associated with the mean effect size under the random-effects model.}
\]

The null hypothesis will be tested under the Z distribution at an alpha level of .05.

**Determining the Power of Detecting Heterogeneity**

Determining the meta-analysis’ ability to detect the heterogeneity of the effect sizes is just as important as its ability to detect effect sizes; detecting heterogeneity determines whether or not there is variation to be explained and, therefore, moderators to be applied. The power to detect heterogeneity under a random effects model is given by:

\[
p = 1 - F(c_\alpha | k - 1; \lambda)
\]

(24)

where,

\[
F(c_\alpha | k - 1; \lambda) = \text{the cumulative distribution function of the non-central chi-square distribution with } k - 1 \text{ degrees of freedom and non-centrality parameter } \lambda, \text{ and where } c_\alpha \text{ is the 100}(1 - \alpha) \text{ percent point of the central chi-square distribution.}
\]

The conventions for \( \lambda \) are as follows:
• Assuming a small degree of heterogeneity, $\lambda = (k - 1)/3$

• Assuming a moderate degree of heterogeneity, $\lambda = 2(k - 1)/3$

• Assuming a large degree of heterogeneity, $\lambda = (k - 1)$

The power of this study to detect a small degree of heterogeneity was calculated to be 0.46, while the power to detect a large degree of heterogeneity was calculated to be 0.98. As such, it is possible that this meta-analysis may miss detecting any heterogeneity if there is only a small amount of variability.

*Finding the Effects of Professional Development by Foci*

Having found the overall mean effect size of professional development on students’ math achievement, it is possible to answer the first research question of whether or not the focus of professional development significantly predicts some of the variation in effect sizes was answered. Prior to describing the method used to answer this question, some notation should be clarified. Let:

- subscript $j$ = the $j$th group or category (for example, each focus will be in their own group or category. Thus, if there are seven foci, then there will be 7 groups).
- subscript $i$ = the $i$th effect size (so, $d_{ji} = $ the $i$th effect size in group $j$).
- $\Delta_j$ = the true population effect in group $j$.
- $d_{j.}$ = the estimated average effect in group $j$.

At this stage, the analysis assumed that the effect sizes are homogeneous within each group. Thus, the random-effects model was represented as:

$$d_{ji} = \Delta_j + e_{ji}$$  \hspace{1cm} (25)
Recall that there was evidence indicating that the effect sizes should differ. Based on the hypothesis presented earlier, it was predicted that these differences were based on the focus of the professional development. The following analysis tested that prediction.

The null hypothesis to be tested is given as:

$$H_0: \Delta_1 = \Delta_2 = \ldots = \Delta_a$$

where each group represents one of the foci, or any other predicted moderator/category.

If the null hypothesis was rejected, then the use of the moderator (such as the focus) as a predictor of the differences is warranted. If the null hypothesis failed to be rejected, then some other predictor is responsible for the differences.

Once the chosen predictor (i.e., the focus) was chosen and applied, the null hypothesis was tested to determine whether or not the focus was a significant predictor of the differences in effect size. To do this, consider the $Q$-statistic used earlier to determine the homogeneity of the effect sizes. This statistic can be represented as the sum of the between-group variation and the unexplained variation such that (Huedo-Medina, et al., 2006):

$$Q = Q_{\text{Between}} + Q_{\text{Within}}$$

where

$$Q_{\text{Between}} = \text{the between-groups variation plus sampling error, and}$$

$$Q_{\text{Within}} = \text{the unexplained between-studies differences plus sampling error.}$$

$Q_{\text{Between}}$ is the homogeneity test for the effect sizes between the groups and can be calculated by:

$$Q_{\text{Between}} = \sum_{j=1}^{p} w_j (d_{j.} - d_{..})^2 \sim \chi^2_{(p-1)}$$

where
where

$$w_{j.} = \frac{1}{\sum_i v_{ji}}$$

(29)

and

$$d_{..} = \frac{\sum_j \sum_i w_{ji} d_{ji}}{\sum_j \sum_i w_{ji}}$$

(30)

and

$$d_{j.} = \frac{\sum_i w_{ji} d_{ji}}{\sum_i w_{ji}}$$

(31)

$Q_{Within}$ is the homogeneity test for the effect sizes within the groups and can be calculated by:

$$Q_{Within} = \sum_j \sum_{i=1}^{p} w_{ji} (d_{ji} - d_{..})^2 = \sum w_{j} \chi^2_{(k-p)}$$

(32)

If the $Q_{Between}$ value is less than the critical value associated with the $\chi^2$ distribution, then the null hypothesis is not rejected and the focus of professional development is not a predictor of the differences of effect sizes. In other words, the three categories of focus of professional development do not lead to significant differences in elementary students’ math achievement. However, if the $Q_{Between}$ value is greater than the critical value, then it would indicate that the focus of professional development does moderate the effects of the professional development. Moreover, the overall effect size for each group could be calculated by applying Equations 16 – 23 for each group.

**Post-hoc Analysis**

Up to this point, the analysis will have indicated whether or not professional development had some overall effect on students’ math achievement. Moreover, the analysis will have identified whether or not the effect of professional development on students’ math achievement differs by foci. The next step was to answer Research
Question 2: What focus of professional development leads to the highest gains in student achievement?

Looking at the means and the 95% CIs will begin to answer this question. Those foci with higher means have greater effects. Non-overlapping confidence intervals give the impression that a pair of foci are significantly different from each other; vice-versa, overlapping confidence intervals for any given pair of foci give the impression that they are not significantly different from each other. However, a more rigorous statistical test can answer this question more concretely. This test uses a series of post hoc contrasts similar to those used in ANOVA, which will indicate whether or not there is a significant difference between the mean effect sizes of the groups. The planned pairwise contrast will be given by (Hedges, 2009).

\[ C = \sum c_j d_j, \quad \text{where } \sum c_j = 0 \quad (33) \]

and has a standard error given by:

\[ SE(C) = \sqrt{\sum c_j^2 v_j}. \quad (34) \]

where

\[ v_j. = \text{the variance of } d_j. \]

A Z-test can be used to test if the pair-wise comparison of the group means are significantly different (e.g.: test the null hypothesis \( H_0: d_1. = d_2. \)). That is:

\[ Z = \frac{\bar{c}}{SE(C)} \quad (35) \]

The null hypothesis will be tested under the Z distribution for a given alpha-level. However, the alpha-level will be determined using the Bonferroni adjustment, which allows for consistent group-wide error. Thus, for each contrast, the alpha level was
determined by the pre-determined group-wide error (i.e., .05) divided by the number of subgroups within a group. For example, assume a contrast is being made determining whether professional development focused on each of the foci had significantly different effects on students’ math achievement. Since there are seven foci, the null hypothesis will be tested under the $Z$ distribution for $\alpha = 0.05/7 = 0.007$.

**The Problem of Multiple Foci per Study**

Analysis using the focus of professional development as a moderator of students’ math achievement was not as easy as it would seem. In an ideal world, each professional development study would limit itself to one focus at a time; and, at most, systematically vary how different foci are blended to create hybrid models. Such an approach would eliminate or limit confusion as to the confounding effects of multiple foci. For example, if a professional development program focuses solely on increasing PCK, then it would be safe to assume that differences in student achievement were due to that focus. In contrast, if a professional development program focused on increasing subject matter knowledge and on cooperative learning, any effects could be due to either focus or a combination of the two foci. Unfortunately for the analysis, there is often more than one focus involved in professional development, as can be seen on Table 1 above.

One method of handling this situation could be simply to ignore the possibility that effects could be due to one focus over another or to a combination of both. This is done by giving each effect size credit for all the foci involved individually. For example, consider a professional development program that focuses on PCK and Equity (such as the study by Cardelle-Elawar, 1990). Each category (PCK & Equity) would be given full
credit for the effect size. In essence, this effect size would be “double-dipping”. Double
dipping is often used and accepted with the warning that confounding effects were not
taken into account (Gleser & Olkin, 2009).

However, another method exists which eliminated the need for “double-dipping”.
According to grounded theory, theoretical frameworks often “require redefinition in order
to fit the realities of qualitative research and the complexities of social phenomena”
(Corbin & Strauss, 1990, p.4). In other words, theory can (some would say “should”) be
adjusted inductively to fit a particular set of data. Referring once again to Table 1, it is
possible to note that the combination of foci of the professional development
interventions fall into following six categories:

a) PCK only
b) SMK and PCK
c) Cooperative Learning
d) PCK and Equity
e) SMK and Efficacy
f) Classroom Management and Shared Vision

If the foci are redefined in this manner, a categorization of the foci of professional
development is created which both eliminates the confounding effects of “double-
dipping” and perfectly fits the set of data used in this meta-analysis. For this reason, it
was these categories of the foci that were introduced as moderators of the effects of
professional development on student gains and not the seven original, individual foci.
CHAPTER 4
RESULTS

Effect Sizes

Description of Effect Sizes

Prior to conducting any analyses, the effect sizes for each study were calculated. Some studies can produce several effect sizes for several measures or populations. For example, study 2 by Cardelle-Elawar (1990) developed an assessment and administered it to a group of students. The results of the test were separated and displayed for males and females (see Table 2 above). Since this study reported the results on two separate populations (i.e., males and females), two effect sizes can be calculated from it. As another example, study 12 by Mason and Good (1993) used only one population, but four separate assessments, each testing computation, conceptual understanding, word problems, or estimation. Therefore, four effect sizes could be calculated for this study. If Mason and Good (1993) had also separated out their results into each of the three grades testes, then 12 effect sizes could have been calculated.

Effect sizes which share a population, such as those computed from Mason and Good’s (1993), create some dependency issues that increase Type II errors. To help limit dependency issues, the inclusion and exclusion of certain effect sizes was considered. Eliminating dependency completely would require that a study be limited to a single measure of the population’s math achievement. In cases where various instruments were used to measure the same attribute (i.e., arithmetic skills or problem solving ability), the instrument that was considered the best by the author was included while the other(s) were excluded. For example, Carpenter et al.’s (1989) study used both a standardized
statewide assessment and a researcher-developed assessment. For purposes of this meta-
analysis, only effect sizes generated from the standardized assessment were used.
However, if the various assessments measured different attributes (i.e., one measures
arithmetic skills and the other measures problem solving abilities) the effect sizes
computed from the various measures were included. Although this creates dependency
issues, the value added by this information was deemed worthwhile and only four of the
studies were affected.

The decision to show preference to state-wide assessments versus researcher
developed measures stemmed from the first purpose of this study: to determine the best
investment in professional development. In the current era of accountability, much value
is being attached to meeting the proficiency standards on each state’s high-stakes exam.
As such, successfully raising achievement scores in these exams is considered a high
priority, and thus would constitute a better investment than raising achievement scores on
researcher-developed tests.

Researcher-developed tests tend to focus on a more specific range of standards
and skills. Because of this, large achievement gains on researcher-developed tests may
not translate to the state-wide exams which test a wider variety of standards and skills.
Moreover, the effect sizes of researcher-developed tests tend to be magnified since the
tests are often aligned directly with the intervention goals which place the control group
at a disadvantage. Thus, using state-wide standardized tests when available should lead
towards more conservative estimates of the effects of professional development on
students’ math achievement.
Table 3 below displays the effect sizes found for the studies included in the meta-
analysis. Within the table, it is noted if dependency issues remain within a certain study.
In other words, whether or not some or all of the effect sizes were derived from the same population.

Two of the studies that met the inclusion criteria did not include enough information to calculate effect sizes. Jagielski (1991) only provided the mean pretest and posttest scores and sample size; the study did not report standard deviations, thus making it impossible to calculate any effect sizes. In an attempt to overcome this impediment, all efforts were made to contact the authors and ask them to provide more details. Of these two studies, only one had contact information that could be located. As of yet, no reply has been received.

The interpretation of these effect sizes is rather loose; however Cohen (1988) established generally accepted guidelines. For social sciences, an effect size of about 0.20 is considered small, 0.50 is moderate, and any effect size greater than or equal to 0.80 is considered large. Looking at the effect sizes listed above, several of them are considered extremely large for educational research. None of the information provided in the study gave the impression that these effect sizes are invalid. Therefore, although the effect sizes are unusual, there was nothing to warrant their exclusion.
<table>
<thead>
<tr>
<th>Study</th>
<th>Measure (Subtests)</th>
<th>Effect Size</th>
<th>Cohen’s $d$ Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burkhouse, Loftus, Sadowski, &amp; Kathy (2003)</td>
<td>PSSA Research Designed</td>
<td>0.54</td>
<td>Moderate</td>
</tr>
<tr>
<td>Cardelle-Elawar (1990)</td>
<td>(Parallel Forms: Males) Research Designed</td>
<td>3.27</td>
<td>Very Large</td>
</tr>
<tr>
<td>Cardelle-Elawar (1990)</td>
<td>(Parallel Forms: Females) Research Designed</td>
<td>2.22</td>
<td>Very Large</td>
</tr>
<tr>
<td>Cardelle-Elawar (1995)</td>
<td>(3rd Grade) Research Designed</td>
<td>1.81</td>
<td>Very Large</td>
</tr>
<tr>
<td>Cardelle-Elawar (1995)</td>
<td>(4th Grade) Research Designed</td>
<td>0.96</td>
<td>Large</td>
</tr>
<tr>
<td>Cardelle-Elawar (1995)</td>
<td>(5th Grade) Research Designed</td>
<td>0.59</td>
<td>Moderate</td>
</tr>
<tr>
<td>Carpenter, Fennema, Peterson, Chiang, &amp; Loef (1989)*</td>
<td>ITBS (Level 7) Computation</td>
<td>0.20</td>
<td>Small</td>
</tr>
<tr>
<td>Carpenter, Fennema, Peterson, Chiang, &amp; Loef (1989)*</td>
<td>ITBS (Level 7) Problem Solving</td>
<td>0.20</td>
<td>Small</td>
</tr>
<tr>
<td>Freiberg, Connell, &amp; Lorentz (2001)</td>
<td>TAAS</td>
<td>0.37</td>
<td>Small/Moderate</td>
</tr>
<tr>
<td>Freiberg, Huziniec, &amp; Templeton (2009)</td>
<td>TAAS Research Designed</td>
<td>0.45</td>
<td>Moderate</td>
</tr>
<tr>
<td>Fuchs, Fuchs, Karns, Hamlett, &amp; Katzaroff (1999)*</td>
<td>Research Designed Conceptual Understanding</td>
<td>1.61</td>
<td>Very Large</td>
</tr>
<tr>
<td>Fuchs, Fuchs, Karns, Hamlett, &amp; Katzaroff (1999)*</td>
<td>Research Designed Computation</td>
<td>1.20</td>
<td>Very Large</td>
</tr>
<tr>
<td>Fuchs, Fuchs, Karns, Hamlett, &amp; Katzaroff (1999)*</td>
<td>Research Designed Problem Solving</td>
<td>1.36</td>
<td>Very Large</td>
</tr>
<tr>
<td>Fuchs, Fuchs, Karns, Hamlett, &amp; Katzaroff (1999)*</td>
<td>Research Designed Communication Derry Township Standardized Test</td>
<td>1.63</td>
<td>Very Large</td>
</tr>
<tr>
<td>Karper &amp; Melnick (1993)*</td>
<td>(3rd Grade Quantitative) Derry Township Standardized Test</td>
<td>-0.09</td>
<td>Small (Neg)</td>
</tr>
<tr>
<td>Karper &amp; Melnick (1993)*</td>
<td>(4th Grade Quantitative) Derry Township Standardized Test</td>
<td>-0.07</td>
<td>Small (Neg)</td>
</tr>
<tr>
<td>Karper &amp; Melnick (1993)*</td>
<td>(5th Grade Quantitative) Derry Township Standardized Test</td>
<td>0.00</td>
<td>None</td>
</tr>
<tr>
<td>Karper &amp; Melnick (1993)*</td>
<td>(3rd Grade Concepts)</td>
<td>0.00</td>
<td>None</td>
</tr>
<tr>
<td>Study</td>
<td>Test</td>
<td>Domain</td>
<td>Effect Size</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>---------------------</td>
<td>---------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Karper &amp; Melnick (1993)*</td>
<td>Derry Township</td>
<td>(4th Grade Concepts)</td>
<td>-0.15</td>
</tr>
<tr>
<td>Karper &amp; Melnick (1993)*</td>
<td>Derry Township</td>
<td>(5th Grade Concepts)</td>
<td>-0.04</td>
</tr>
<tr>
<td>Karper &amp; Melnick (1993)*</td>
<td>Derry Township</td>
<td>(3rd Grade Computation)</td>
<td>0.01</td>
</tr>
<tr>
<td>Karper &amp; Melnick (1993)*</td>
<td>Derry Township</td>
<td>(4th Grade Computation)</td>
<td>-0.08</td>
</tr>
<tr>
<td>Karper &amp; Melnick (1993)*</td>
<td>Derry Township</td>
<td>(5th Grade Computation)</td>
<td>-0.07</td>
</tr>
<tr>
<td>Martin (2006)</td>
<td>Research Designed</td>
<td>(4th Grade; Year 1)</td>
<td>0.28</td>
</tr>
<tr>
<td>Martin (2006)</td>
<td>Research Designed</td>
<td>(4th Grade; Year 2)</td>
<td>0.87</td>
</tr>
<tr>
<td>Martin (2006)</td>
<td>Research Designed</td>
<td>(4th Grade; Year 3)</td>
<td>0.44</td>
</tr>
<tr>
<td>Martin (2006)</td>
<td>Research Designed</td>
<td>(5th Grade; Year 1)</td>
<td>0.24</td>
</tr>
<tr>
<td>Martin (2006)</td>
<td>Research Designed</td>
<td>(5th Grade; Year 2)</td>
<td>0.27</td>
</tr>
<tr>
<td>Martin (2006)</td>
<td>Research Designed</td>
<td>(5th Grade; Year 3)</td>
<td>0.45</td>
</tr>
<tr>
<td>Mason &amp; Good (1993)*</td>
<td>Research Designed</td>
<td>(Computation)</td>
<td>0.33</td>
</tr>
<tr>
<td>Mason &amp; Good (1993)*</td>
<td>Research Designed</td>
<td>(Conceptual Understanding)</td>
<td>0.15</td>
</tr>
<tr>
<td>Mason &amp; Good (1993)*</td>
<td>Research Designed</td>
<td>(Problem Solving)</td>
<td>0.13</td>
</tr>
<tr>
<td>Mason &amp; Good (1993)*</td>
<td>Research Designed</td>
<td>(Estimation)</td>
<td>0.29</td>
</tr>
<tr>
<td>Pace (2008)</td>
<td>CSAP</td>
<td>(3rd Grade)</td>
<td>0.02</td>
</tr>
<tr>
<td>Pace (2008)</td>
<td>CSAP</td>
<td>(4th Grade)</td>
<td>0.08</td>
</tr>
<tr>
<td>Pace (2008)</td>
<td>CSAP</td>
<td>(5th Grade)</td>
<td>-0.15</td>
</tr>
<tr>
<td>Ross, Bruce, &amp; Hogaboam-Gray (2006)</td>
<td>EQAO</td>
<td></td>
<td>0.11</td>
</tr>
<tr>
<td>Sloan (1993)</td>
<td>CTBS</td>
<td></td>
<td>0.08</td>
</tr>
</tbody>
</table>
Homogeneity of the Effect Sizes

Looking at this CI Plot below (Figure 5), it seems that the effect sizes are not homogeneous. As a general rule of thumb, if a horizontal line can be drawn that crosses all the confidence intervals, then there is probably no statistically significant difference in the effect sizes. This does not seem to be the case.

The $Q$-statistic indicated that the effect sizes included in this analysis are not homogeneous; $Q(41) = 688.36, p < .05$. This has two consequences. First, overall effect sizes were not from the same population, which supports using a random effects model. Second, it is now possible to explore sources of variation in the effect sizes. In other words, it is now possible to test if the focus of the professional development is a significant predictor of the variation of the effect sizes (Research Question 1). Both of these consequences are discussed in greater detail in the following sections.
Figure 5: Effect sizes with 95% confidence intervals

Note: The study label x.y refers to study x, effect size y (e.g.: 10.2 is the second effect size for study 10).

**Overall Effect of Professional Development on Student Gains**

Applying the above analysis for determining the overall mean effect size under the random effects model, the results show that the overall effect size of professional development on students’ math achievement is 0.57(SE = 0.13). Both the 95% CI of 0.37 to 0.77 and the Z-test of the overall mean ($Z = 4.38, p < .05$), indicate that the effect size is significantly different from zero.
RQ1: Does the Focus of Professional Development Predict Some Variation of Effect Sizes?

The results from this analysis indicated that using the foci-groupings as a moderator is a significant predictor of the variation of the effect sizes;

\[ Q_{\text{between}}(5) = 154.03, \ p < .05 \]. In other words, the focus of professional development does significantly predict the variation of effects of professional development on student gains in mathematics. However, \( Q_{\text{within}} \), indicated that there is more variability left to explain;

\[ Q_{\text{within}}(36) = 534.33, \ p < .05 \]. Thus, the above process was repeated within each foci-group. For some foci, \( Q_{wj} \) was found to not be significant. In these cases, a fixed-effects weighting (i.e., the inverse variance) was applied. The results from this analysis are listed in Table 4. Figure 6 graphically represent the overall means and 95% CI of the overall means by focus.

**Table 4: Summary of Results**

<table>
<thead>
<tr>
<th>Focus Category</th>
<th># of Studies</th>
<th># of Effect Sizes</th>
<th>Mean</th>
<th>SE</th>
<th>Z</th>
<th>95% CI</th>
<th>( Q_{wj} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall, ( Q_{\text{Total}}(41) = 688.36^* )</td>
<td>15</td>
<td>42</td>
<td>0.57</td>
<td>0.13</td>
<td>4.38*</td>
<td>[0.32, 0.82]</td>
<td>-</td>
</tr>
<tr>
<td>Focus of PD: ( Q_{\text{between}}(5) = 154.03^<em>; Q_{\text{within}}(36) = 534.33^</em> )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMK &amp; PCK</td>
<td>2</td>
<td>9</td>
<td>0.26</td>
<td>0.10</td>
<td>2.60*</td>
<td>[0.07, 0.45]</td>
<td>84.49^*</td>
</tr>
<tr>
<td>PCK Only</td>
<td>7</td>
<td>15</td>
<td>1.05</td>
<td>0.20</td>
<td>5.25*</td>
<td>[0.66, 1.43]</td>
<td>425.16^*</td>
</tr>
<tr>
<td>Coop Learning Only</td>
<td>2</td>
<td>13</td>
<td>0.19</td>
<td>0.03</td>
<td>6.33*</td>
<td>[0.14, 0.24]</td>
<td>21.42^*</td>
</tr>
<tr>
<td>PCK &amp; Equity</td>
<td>1</td>
<td>2</td>
<td>2.65</td>
<td>0.31</td>
<td>8.55*</td>
<td>[2.04, 3.25]</td>
<td>2.79</td>
</tr>
<tr>
<td>SMK &amp; Math Efficacy</td>
<td>1</td>
<td>1</td>
<td>0.25</td>
<td>0.04</td>
<td>6.25*</td>
<td>[0.17, 0.33]</td>
<td>N/A</td>
</tr>
<tr>
<td>Classroom Mgmt &amp; Shared Vision</td>
<td>2</td>
<td>2</td>
<td>0.42</td>
<td>0.04</td>
<td>10.50*</td>
<td>[0.34, 0.50]</td>
<td>0.47</td>
</tr>
</tbody>
</table>

* Significant at \( p < .05 \)
Figure 6: Overall means and 95% CI for each Focus
RQ2: Which Focus Leads to the Greatest Student Gains?

A series of contrasts were made to test for significant differences in the effects of the foci on student achievement. The results of the post-hoc analyses are listed in Table 6.

Table 5: Post-hoc contrasts

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Z-score</th>
<th>p-value</th>
<th>Significantly Different?*</th>
<th>Larger Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMK &amp; PCK vs. PCK Only</td>
<td>11.39</td>
<td>&lt;.001</td>
<td>Yes</td>
<td>PCK Only</td>
</tr>
<tr>
<td>PCK &amp; Equity vs. SMK &amp; Math Efficacy</td>
<td>8.84</td>
<td>&lt;.001</td>
<td>Yes</td>
<td>Equity</td>
</tr>
<tr>
<td>SMK &amp; PCK vs. PCK &amp; Equity</td>
<td>7.38</td>
<td>&lt;.001</td>
<td>Yes</td>
<td>Equity</td>
</tr>
<tr>
<td>PCK Only vs. PCK &amp; Equity</td>
<td>4.35</td>
<td>&lt;.001</td>
<td>Yes</td>
<td>PCK Only</td>
</tr>
<tr>
<td>PCK Only vs. Coop Learning</td>
<td>4.20</td>
<td>&lt;.001</td>
<td>Yes</td>
<td>PCK Only</td>
</tr>
<tr>
<td>PCK Only vs. SMK &amp; Efficacy</td>
<td>2.85</td>
<td>0.004</td>
<td>Yes</td>
<td>PCK Only</td>
</tr>
<tr>
<td>PCK Only vs. Classroom Mgmt &amp; Shared Vision</td>
<td>3.11</td>
<td>0.002</td>
<td>Yes</td>
<td>PCK Only</td>
</tr>
<tr>
<td>PCK &amp; Equity vs. Coop Learning</td>
<td>7.81</td>
<td>&lt;.001</td>
<td>Yes</td>
<td>PCK &amp; Equity</td>
</tr>
<tr>
<td>PCK &amp; Equity vs. SMK &amp; Efficacy</td>
<td>4.51</td>
<td>&lt;.001</td>
<td>Yes</td>
<td>Equity</td>
</tr>
<tr>
<td>PCK &amp; Equity vs. Classroom Mgmt &amp; Shared Vision</td>
<td>7.12</td>
<td>&lt;.001</td>
<td>Yes</td>
<td>Equity</td>
</tr>
</tbody>
</table>

*Significant p-value based on .05 divided by the total number comparison groups used in post-hoc test; p < .005
CHAPTER 5
CONCLUSIONS AND DISCUSSION

This study sought to answer two questions: does the focus of professional development significantly explain some of the variation of the effects of professional development on student gains in mathematics; and if so, what focus of professional development leads to the greatest student gains in mathematics. However, this study provided information beyond the breadth of those two questions. For example, whereas the previous meta-analyses (Blank & de las Alas, 2009; Scher & O’Reilly, 2009; Slavin & Lake, 2008) have shown that professional development has a small positive effect on student achievement, the results of this meta-analysis indicate that the effect may possibly be larger than previously found. According to these results, professional development has a moderate effect on student gains in mathematics.

It is possible that this result is over-inflated due to what is referred to as the “file drawer” problem. This problem assumes that the easiest studies to find while conducting the literature search for a meta-analysis are those studies which have been published (Rosenthal, 1979). It also assumes that the articles which are most likely to be published are those which show the desired effects. In other words, studies which indicate undesirable or no effects are less likely to be published, thus remain in the author’s file drawer. As a result, the literature included in meta-analyses is mostly limited to those studies which have desirable results and, therefore, bias the results of the meta-analysis. To demonstrate this, a funnel plot, such as shown in Figure 7 plots the effect sizes versus the sample sizes for the studies included in this study.
Figure 7: Funnel plot demonstrating the “file-drawer problem”

An expected distribution for this type of plot would look like an upside-down funnel (Egger, Smith, Schneider, & Minder, 1997). That is, there would be a peak with a somewhat equal distribution on either side of that peak. The data in Figure 10 lie primarily on the right side of the peak, meaning that the overall effect size indicated by the study is biased towards the right; the overall effect sizes are probably too large. This is supported by Egger’s Regression Test of Symmetry which statistically tests whether or not the effect sizes are distributed symmetrically around a mean; \( t = 3.87, p < .05 \). In an effort to overcome this limitation strong efforts have been made to identify and include unpublished studies. A variety of strategies were employed, including contacting colleagues and other professionals in the field to see if they know of anyone who may
have unpublished work on this topic and procuring unpublished studies that have been cited in other studies dealing with professional development.

It is important to keep in mind, however, that this is a limitation that all meta-analyses in education research experience. Logically then, the previous meta-analyses which indicated small effect sizes of professional development on student gains in mathematics also were subject to inflated results. Therefore, the results of this meta-analysis still indicated that the effect of professional development was larger than previously thought.

RQ1: Conclusion and Discussion

The results demonstrate that the focus of professional development does significantly explain some of the variation of the effects of professional development on student gains in mathematics. This validates that attention should be paid to the substantive content of professional development beyond the structural characteristics. It also validates the use of the six foci (i.e., PCK & SMK, PCK Only, etc.) derived from the redefinition of the a priori categories. Moreover, this informs policy-makers and practitioners that attention should be given to the focus of any professional development intervention in order to make better investments for their teachers and students.

Just as this study provides information about the size of the effect of professional development on students’ math achievement, it also provided unintended information about the state of the literature in this field. Using the categories developed by this study required the 15 studies included in the meta-analysis to be grouped into six separate categories. Table 4 above highlights how, as opposed to the seven studies which focused
on PCK only, most of the foci categories included only one or two studies. This does not mean that the categorization schema should not have been used for this analysis; the categories of foci used in this analysis were based on theory and the literature. However, the current body of literature is insufficiently large to realize the promise of possibilities created by this categorization schema. Beyond issuing some caution as to the results found in this study, this information highlights that more research needs to be done in order to accurately flesh out these gaps in the literature. Doing so would allow this category system to be appropriately applied to determine the best investment in professional development by means other than structural components, as well as identify what knowledge and beliefs teachers need (or need more of) in order to teach effectively.

Viewing the larger picture, the majority of the studies identified for inclusion focused on changing teachers’ knowledge of SMK and/or PCK while many fewer looked to change teachers’ knowledge of pedagogy (e.g.: cooperative learning and classroom management), shared vision, sense of self as a math teacher, or knowledge of equity. Thus, the question arises as to why this is so. Recall that the inclusion criteria for this study required that the professional development intervention be tied to some student achievement measures. Do many studies that focus more on content other than SMK and PCK exist, but do not connect the intervention to student achievement? Or, are these studies not being conducted? If they are not being connected to student achievement, why not? If the professional development interventions are not giving the same emphasis to these other foci, why not? These are questions that, if answered, may speak to the values being held in education today.
RQ2: Conclusion and Discussion

According to Table 5, professional development with a substantive content focus on both PCK and Equity leads to the highest student gains in mathematics, followed by a focus on PCK only. Thus, if these results were to be trusted, the best investment in professional development would be in professional development having the structural characteristics supported by Porter et al. (2003) and focused on changing teachers’ knowledge of PCK and Equity. However, it is necessary to keep in mind that, because of the sparseness of the literature, the PCK and Equity category was comprised of only one study. As such, it is difficult to determine the soundness of this particular result.

On the one hand, it is difficult to place much emphasis on results based on a single study. On the other hand, there is nothing to indicate that this study produced unique results. In fact, studies such as The Algebra Project (Moses, Kamii, Swap, & Howard, 1989) and QUASAR (Silver & Stein, 1996) are examples of successful interventions that have explicitly focused on PCK for math and Equity. Both of these interventions has demonstrated significant, and often large, effects on students’ math achievement, but they were excluded from this meta-analysis because they either fell out of the specified date range or targeting middle school teachers. Nonetheless, this particular result of the large effect on student gains of professional development focused on both PCK and Equity must be used with caution since there is an apparent lack of statistical power to support it. There is, however, theoretical support for the large gains indicated by these results.
Theoretical Support of PCK and Equity.

As discussed previously in Chapter 2, effective teachers need to be informed of what their students know, what student misconceptions are being held, and how to build and reform these misconceptions. In other words, effective teachers need PCK. Moreover, PCK enables teachers to select valuable activities, have a wide range of explanations of concepts, ask helpful questions, and assess student understanding (Hill & Lubienski, 2007; Shulman, 1986; Ball & McDiarmid, n.d.). Implicit in the PCK literature is the assumption that this conception of PCK applies equally to all teachers of all students – including teachers who encounter diverse student populations in their classrooms. However, PCK literature does not necessarily discuss that student misconceptions, what constitutes valuable activities, etc. may differ based upon the student population. This can be seen in the original NCTM Standards documents; both the Curriculum and Evaluation Standards for School Mathematics (NCTM, 1989) and the Professional Standards for Teaching Mathematics (NCTM, 1991) have been criticized for giving the impression that all students could be equitably served through a single pedagogy (Lubienski, 2002). The more recent Principles and Standards for School Mathematics (NCTM, 2000) begins to address the issue of equitable pedagogy by creating an Equity Principle which states that “equity does not mean that every student should receive identical instruction; instead, it demands that reasonable and appropriate accommodations be made as needed to promote access and attainment for all students” (p. 12). Yet, this document is still not clear or specific as to “reasonable and appropriate accommodations” might be for linguistically and culturally minority students. This strand of research has, in general, been the realm of multicultural educationalists.
Teachers who are successful with low-SES, minority students, or bilingual students may have a different set of teaching practices and values than teachers who are unsuccessful with these students or those teachers placed in White, middle-class classrooms (Boaler, 2002; Ladson-Billings & Darling-Hammond, 2000). Some researchers, policy-makers, and teachers have dichotomized the successful versus the unsuccessful teachers of low-SES, minority, and bilingual students into those who teach traditional mathematics and those who teach reform mathematics; they further argue that low-SES and minority students do better in traditional math classrooms (Boaler, 2002; Lubienski, 2002). Most of the PCK being taught in professional development is based on reform mathematics.

In traditional mathematics, teachers demonstrate a problem and then expect students to solve similar problems. In contrast, in reform mathematics, students construct knowledge through discussion and open-ended word problems set in “real-world” contexts. Whereas a teacher of traditional mathematics would be satisfied with a correct answer, a teacher of reform mathematics would require students to explain and justify their answer, regardless of whether or not it was correct.

Some research does support the claim that minority, low-SES, and bilingual students do better in traditional classrooms. In her study, Lubienski (2002) found that low-SES students were less self-assured and less successful than the middle class students in a reform classroom. Moreover, she saw inequitable participation of these students in whole-class discussions and contextualized problems. These low-SES students placed the blame squarely on the shoulders of the open-ended nature of the problems; they wanted clear and explicit facts and rules to follow. However, Lubienski
(2002) and Boaler (2002) warn that these results must be carefully reported and considered. They argue for the need to be careful about not perpetuating a pedagogy of poverty, in which some students are viewed as less capable of demanding work.

“Educators must understand that the needs of different groups of students not to develop negative ideas about the students’ mathematical potential, but to become aware of the ways in which schools can serve all students” (Boaler, 2002, p. 241). To support this warning, Boaler (2002) and Lubienski (2002) each discuss how some reform approaches promote equity while others fail to do so. The trick, each argues, is in identifying how these approaches differed. Accordingly, some reform practices that are productive in promoting equity should be the focus of teacher learning strategies that will support these practices.

Multiculturalists have a variety of methods describing and explaining the need for different approaches to teaching and the need for more specified/specialized PCK. Delpit (1988), for example, stated that schools represent a culture of power. That is, the norms and values held within schools are that of the culture in power. In the case of the United States, the culture of power is that of the White, middle-class. Therefore, any student who does not come from the culture of power is immediately disadvantaged and must be made explicitly aware of the norms and values in order to have a chance at being successful.

Not dissimilarly, Bourdieu (1973) argued that each student arrives in school with his or her own set of cultural norms and values, which he referred to as cultural capital. However, instead of necessarily promoting the explicit enculturation of students into the culture of power, Bourdieu holds that teachers must be made aware of their students’ cultural capital and incorporate it into their practices and pedagogies. Building on this
idea, Nelson-Barber and Estrin (1995) argued that “[a]ll cultures generate mathematical and scientific knowledge, but that knowledge need not look the same from one group to another” (p. 176). Likewise, people of all cultures learn, but the way they learn need not look the same from one group to another.

Considering Delpit’s and Bourdieu’s ideas of a culture of power and cultural capital does not lead to believing different PCK is necessary for different populations of students. Instead, it means that teachers of low-SES and minority students need to go beyond general PCK; they should include strategies and knowledge on how to help prepare these students to learn in reform mathematics classrooms. This, then, could explain the high effect of this combined focus of professional development on students’ math achievement as opposed to focusing on PCK only.

Effective practices, strategies, and types of PCK, all of which also focus on equity, are beginning to come to light. Boaler (2002) and Lubienski (2002) have both identified similar strategies in the repertoires of effective teachers of low-SES and minority students in reform mathematics classrooms. These teachers choose valuable activities, ask helpful questions, and identify student misconceptions, but do so with the students’ home culture in mind. For example, African American and low-SES children are said to come from cultures that expect rules to be communicated explicitly while White, middle class students often experience a culture in which they need to interpret their parents’ indirect statements (Delpit, 1983). Thus, effective teachers of African American and low-SES students spend significant amounts of time introducing activities through discussion in order to ensure that students understand the expectations and what is being asked, as well as to judge what level of support the students might need.
Other research has demonstrated that White, middle-class students recognize how to limit and contextualize “real-world problems” for the purpose it was asked. In contrast, low-SES students apply the context of “real-world problems” to their experiences in the “real-world”. For instance, when determining whether a weekly bus pass or a daily bus pass is more economical, White, middle-class students will refer only to the price of each while low-SES students begin to discuss how often they themselves use the bus, whether or not the pass is transferable, etc., applying their true concerns and experiences (Boaler, 2002). Thus, teachers with this specialized PCK will either accept the answers based on the various contexts supplied by the students, or explicitly teach the students how to re-contextualize these types of problems for the purpose of schooling.

In general, these effective practices that combine PCK with Equity recognize that students come to school with different cultural capital into an environment in which a culture of power dominates. Thus, these teachers use this knowledge to synchronize the disparate cultures and choose both cultural and content relevant pedagogy, and have the potential to substantially affect student performance.

*Professional Development Focused on PCK Only as the Best Investment*

Given that the results of focusing on PCK and Equity are on shaky ground empirically, if not theoretically, professional development focused on PCK only could be viewed as the best investment until more research is conducted to flesh out the other categories. Not only does this consideration have a stronger empirical backing for now, but it has theoretical backing as well.
The emphasis placed on PCK is based on the assumption that being an expert of mathematics is often not sufficient for the teaching of mathematics (Ball & Bass, 2002). Shulman (1986), in fact, defined PCK as:

the most useful form of [content] representation…, the most powerful of analogies, illustrations, examples, explanations, and demonstrations – in a word, the ways of representing and formulating the subject that makes it comprehensible for others (p. 9).

Moreover, Shulman argued that PCK is what would discriminate between a mathematician and a math educator. This is partially because PCK integrates all other categories of teacher knowledge (Gess-Newsome, 2001). That is, it combines and exists within the boundaries of subject matter knowledge, pedagogical knowledge, and beliefs (Ball & Bass, 2002). Whereas a mathematician need only know mathematics and a teacher of any subject has general pedagogical knowledge (e.g.: classroom management and cooperative learning strategies), a math educator must be able to apply knowledge of mathematics to pedagogy and knowledge of pedagogy to mathematics. It is this area of knowledge of teaching in combination with knowledge of content from which PCK seems to gain its strength to affect student learning.

Grossman’s (1990) four categories of PCK and Ball, Thames, and Phelps’ (2008) three categories of PCK help elucidate the marriage between content knowledge and pedagogical knowledge required for PCK. According to Grossman (1990), PCK is composed of: (a) a teacher’s belief about the subject and the purposes for teaching it, (b) a teacher’s knowledge of what students understand or may potentially misunderstand about a particular subject, (c) a teacher’s knowledge of the curriculum, and (d) a teacher’s knowledge of the various representations and strategies for teaching a particular subject. Each of these components is comprised of a combination of content and pedagogy.
Similarly, Ball, Thames, and Phelps (2008) stated that PCK is composed of: (a) knowledge of content and students, (b) knowledge of content and teaching, and (c) knowledge of content and curriculum.

Teachers perform a wide variety of tasks as part of their practice. This includes designing quizzes, selecting and modifying activities, posing questions, managing discussions, keeping track of students’ learning, determining the validity of student answers, and deciding the classroom objectives (Ball & Bass, 2002). These tasks require knowledge of math and how it interacts with teaching, learning, and the curriculum (Ball & Bass, 2002; Putnam & Borko, 1996). In other words, these tasks require PCK. PCK is used as a conceptual tool for enhancing instructional decision-making. This tool has been found to be especially important for teachers who emphasize teaching for understanding instead of memorization (Putnam & Borko, 1996). Therefore, it would seem that increasing teachers’ PCK in math should be of particular importance for increasing students’ math understanding and achievement.

Ball and Bass (2002) argued that all teachers have some level of mathematical knowledge. However,

[what many teachers lack is mathematical knowledge that is useful and usable for teaching. Of course, some teachers do learn some mathematics in this way from their teaching, from using curricular materials thoughtfully and by analyzing student work. However, many do not. Inadequate opportunities exist for teachers to learn mathematics in ways that prepare them for the work [of teaching], and few curriculum materials effectively realize their potential to provide mathematical guidance and learning opportunities for these teachers (Ball & Bass, 2002, p. 13).

In their results of a research synthesis concerning teacher knowledge, Putnam and Borko (1996) also stated that teachers, especially novice teachers, tended to have underdeveloped PCK. Professional development focused on increasing teachers' PCK in
mathematics would therefore provide teachers with something many of them lack and which has been deemed as important (the most important, according to Shulman) for teachers to effectively teach for understanding.
CHAPTER 6
IMPLICATIONS AND FUTURE RESEARCH

With the majority of American elementary school students failing to meet basic proficiency standards in mathematics, there is a danger that policy makers might rush to find a quick fix. Although this is understandable, a quick fix may not be the best investment of schools’ money or teachers’ time. This meta-analysis has demonstrated, as have meta-analyses before, that professional development, in general, will have a positive effect on student’s math achievement. It has also provided some evidence that the focus of the professional development is an important consideration; some foci have greater effects than others. Lastly, this study has offered some indication as to what focus on professional development would represent the best investment. However, in addition to the evidence and indications provided, this study has also made clear that there is still much work to be done before a true “best investment” can be solidly identified.

Creating professional development interventions which focus on a combination of PCK and Equity for the explicit purpose of making the best investment would be a quick fix. However, because this result is based solely on one study, this quick fix may not necessarily be a good fix, regardless of the results of this meta-analysis. On the other hand, creating professional development interventions with these foci in order to replicate, confirm, or disprove these results would be a strong step forward in the process of finding the best investment. More than the one identified study should be included as part of this decision-making process. What these results truly depict is that there may be something powerful with professional development having this dual focus, and that these results should continue to be investigated.
This recommendation to investigate the potential of professional development which focuses on both PCK and Equity simultaneously is something that has not previously been given much emphasis in the existing literature. Studies on the effectiveness of professional development have either not taken equity as a substantive focus into consideration (Loucks-Horsely & Matsumoto, 1999; Porter et al., 2003), or have isolated it as an entirely separate focus (Cohen & Hill, 2000). For example, when Cohen and Hill (2000) measured the effectiveness of professional development with different foci to positively change teacher practices, they concluded that professional development focused on special topics such as equity had no effect while a focus on teaching math led to significantly positive changes in teacher practices. Perhaps Cohen and Hill’s (2000) conclusion is one of the reasons that no studies were found by this meta-analysis that focused solely on equity. Based on the results of this meta-analysis, however, it seems as if there may be some synergy created when these two foci are applied together that has not been previously considered and should be more thoroughly explored.

Just as other studies have failed to consider the possible synergy created when professional development focuses simultaneously on PCK and Equity, it is possible that studies have also failed to consider synergy that may be created when professional development is combined with technology or curriculum interventions. For the purposes of this study, those professional development interventions that also included technology or curriculum interventions were excluded so as not to confound the effects of the foci of professional development on students’ math achievement. However, it may be worthwhile in the future to examine how professional development in isolation compares
to professional development combined with a technology or curricular intervention. One of these combinations may also result in a synergistic relationship.

Lastly, this study solely investigated the effects of professional development on student achievement. This was based on the current political climate which emphasizes accountability and high-stakes testing, thus leading to the assumption that the best investment leads to increased achievement gains. However, Figure 1 illustrated that professional development also has the capability of impacting student attitudes and beliefs. These affective student attributes also have value. For instance, when the current White, middle-class male composition of the STEM (Science, Technology, Engineering, and Mathematics) fields is considered, changing minority and female students’ attitudes and beliefs about math may play a vital role in changing the future composition of these fields. Interventions such as Multiplying Options and Subtracting Bias (Fennema, Wolleat, Pedro, & Becker, 1981) have proven that this is possible. As such, future studies may consider how to best invest in professional development that effectively changes these affective student outcomes instead of student achievement.
REFERENCES

References marked with an asterisk (*) indicate studies included in the meta-analysis.


## Appendix A: Coding Sheet

<table>
<thead>
<tr>
<th>Construct</th>
<th>Variable</th>
<th>Variable Label</th>
<th>Variable Name</th>
<th>Define Values</th>
<th>Page #</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>**Article</td>
<td>Document ID</td>
<td>Document ID</td>
<td>Doc_ID</td>
<td>Numerical</td>
<td>N/A</td>
<td>Previously Assigned</td>
</tr>
<tr>
<td>Characteristics</td>
<td>Author(s)</td>
<td>Author(s)</td>
<td>Authors</td>
<td>String</td>
<td>N/A</td>
<td>First Initial, Middle Initial, Last Name</td>
</tr>
<tr>
<td></td>
<td>Publication Date</td>
<td>Publication Date</td>
<td>Pub_Date</td>
<td>Numerical</td>
<td>N/A</td>
<td>Year is sufficient</td>
</tr>
<tr>
<td>Source</td>
<td>Publication Source</td>
<td>Source</td>
<td>Source</td>
<td>1) Journal</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2) Book Chapter</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3) Dissertation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4) Conference Center</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5) Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6) Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Source</td>
<td>Specify which &quot;other&quot; source</td>
<td>Source_Other</td>
<td>String</td>
<td>N/A</td>
<td></td>
<td>Only answer this if Source = 6</td>
</tr>
<tr>
<td>Sponsor</td>
<td>Was the study sponsored?</td>
<td>Sponsor</td>
<td>1) Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2) No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3) Not indicated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sponsor Name</td>
<td>Specify who the sponsor</td>
<td>Sponsor_Name</td>
<td>String</td>
<td></td>
<td></td>
<td>Only answer if Sponsor = 1</td>
</tr>
<tr>
<td>Relevance</td>
<td>Relevant Date</td>
<td>Rele_Date</td>
<td>1) Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2) No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Elementary School</td>
<td>Did the study include elementary school teachers/students?</td>
<td>Rele_Elem</td>
<td>1) Yes</td>
<td>2) No</td>
<td>N/A</td>
<td>Elementary School includes grades 1 thru 6, unless 6th grade is part of a middle school</td>
</tr>
<tr>
<td>--------------------</td>
<td>-------------------------------------------------------------</td>
<td>----------</td>
<td>--------</td>
<td>--------</td>
<td>-----</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Quantitative Study</td>
<td>Is this an empirical quantitative study?</td>
<td>Rele_Quant</td>
<td>1) Yes</td>
<td>2) No</td>
<td>N/A</td>
<td>Excluding research syntheses and meta-analysis</td>
</tr>
<tr>
<td>In-service PD</td>
<td>Was the PD given to in-service teacherse?</td>
<td>Rele_InServe</td>
<td>1) Yes</td>
<td>2) No</td>
<td></td>
<td>Exclude PD focused on curriculum, teaching materials, or part of a Comprehensive School Reform</td>
</tr>
<tr>
<td>Math Outcomes</td>
<td>Are there student achievement outcomes in math?</td>
<td>Rele_math</td>
<td>1) Yes</td>
<td>2) No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effect of PD</td>
<td>Does the study provide at least one student achievement outcome as an effect of the PD?</td>
<td>Rele_effect</td>
<td>1) Yes</td>
<td>2) No</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Study Design | Specify type of study design | Rele_Design | 1) Randomized Control Trial (RCT)  
2) Quasi-experimental Design (QED)  
3) Single-subject Design (SSD)  
4) Regression Discontinuity Design (RDD)  
5) Other |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Other Design</td>
<td>Specify the type of &quot;other&quot; design</td>
<td>Design_Other</td>
<td>String</td>
</tr>
<tr>
<td>Duration</td>
<td>The duration of the PD intervention was:</td>
<td>Rele_Duration</td>
<td>String</td>
</tr>
</tbody>
</table>
| Relevance | Is the study relevant? | Relevance | 1) Yes  
2) No | N/A | Answer "NO" if any of the Relevance Constructs is "NO". IF NOT RELEVANT: STOP CODING! |
<p>| Quality of Study | Student Outcomes | St_Outcomes | String | If available: include name, type, and any psychometric information |</p>
<table>
<thead>
<tr>
<th>Outcome Quality</th>
<th>Does at least one of the student achievement outcome measures have face validity, report reliability, or is a standardized test?</th>
<th>Qual_Outcome</th>
<th>1) Yes</th>
<th>2) No</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome Quality Description</td>
<td>Specify what quality information is reported for the student outcome measures</td>
<td>Qual_Out_Desc</td>
<td>String</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCT Procedure</td>
<td>Did the authors provide details of the randomization Procedure</td>
<td>RCT_Procedure</td>
<td>1) Yes</td>
<td>2) No</td>
<td></td>
</tr>
<tr>
<td>RCT Problem Free</td>
<td>Is the study free of any problems with randomization?</td>
<td>RCT_Free</td>
<td>1) Yes</td>
<td>2) No</td>
<td></td>
</tr>
</tbody>
</table>

This can include descriptions of face validity, reliability information, if it is a standardized test, etc.

ONLY ANSWER IF STUDY IS AN RCT

Problems can include subjects being switched after assignment, uneven attrition, intervention disruption, etc.
<table>
<thead>
<tr>
<th>RCT Problem Description</th>
<th>Specify what kind, if any, problems were identified with the RCT Procedure</th>
<th>RCT_Free_Desc</th>
<th>String</th>
<th>See RCT_Free notes for possible problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Method</td>
<td>Which of the following were used in order to control for group differences?</td>
<td>QED_Control</td>
<td>1) group matching 2) statistical control 3) not informed 4) other</td>
<td>ONLY ANSWER IF STUDY IS A QED</td>
</tr>
<tr>
<td>Control Method Other</td>
<td>What was the &quot;other&quot; group difference control method?</td>
<td>Control_Other</td>
<td>String</td>
<td>Only answer if QED_Control = 4</td>
</tr>
<tr>
<td>Control Described</td>
<td>Do the authors describe the matching or statistical control procedures?</td>
<td>Control_Desc</td>
<td>1) Yes 2) No</td>
<td>ONLY ANSWER IF STUDY IS A QED</td>
</tr>
<tr>
<td>QED Baseline</td>
<td>Is it apparent that baseline characteristics were appropriately matched or accounted for with statistical adjustments?</td>
<td>QED_Baseline</td>
<td>1) Yes 2) No</td>
<td>ONLY ANSWER IF STUDY IS A QED. As described by the authors or by looking at the control and treatment characteristic tables</td>
</tr>
<tr>
<td>Category</td>
<td>Question</td>
<td>Answer 1</td>
<td>Answer 2</td>
<td>Additional Information</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Stability</td>
<td>Was there stability in the baseline condition before treatment was introduced?</td>
<td>SSD_\text{Stability}</td>
<td>1) Yes</td>
<td>Only answer if study is an SSD. If things were not stable prior to treatment, then change may not be due to PD</td>
</tr>
<tr>
<td>SSD Variable</td>
<td>Was only one variable changed during the treatment condition?</td>
<td>SSD_\text{Variable}</td>
<td>1) Yes</td>
<td>Only answer if study is an SSD. If more than one variable changed, then change cannot be attributed to PD</td>
</tr>
<tr>
<td>Cut-off Criteria</td>
<td>Was the cut-off criteria described, based from a single continuous pretest distribution, and followed without exception?</td>
<td>RDD_\text{Criteria}</td>
<td>1) Yes</td>
<td>Only answer if study is an RDD</td>
</tr>
<tr>
<td>RDD Distribution</td>
<td>Does the pre-post distribution follow a polynomial function?</td>
<td>PDD_\text{Distribution}</td>
<td>1) Yes</td>
<td>Only answer if study is an RDD</td>
</tr>
</tbody>
</table>
| Uniform Implementation | Is the PD uniformly implemented to all recipients under the same conditions? | Uniform_PD | 1) Yes  
2) No | Same conditions should include duration, frequency, kind, etc. |
|------------------------|--------------------------------------------------------------------------------|------------|--------------|-----------------------------------------------------------|
| Confound Free          | Is the study free of any teacher-intervention confound problems?                | Confound_Free | 1) Yes  
2) No | These should be specified by the author. |
| Confound Description   | If there was a confound, please describe it.                                   | Confound_Desc | String | Only answer if Confound_Free = 2 |
| Attrition              | Is the study free from any overall of differential attrition problems?          | Attrition   | 1) Yes  
2) No | This can be author or coder identified. Includes extreme attrition overall, more attrition for one group, etc. |
| Attrition Description  | What type of attrition problem was identified?                                | Attrition_Desc | String | Only answer if Attrition = 2 |
| Finding of Effect Sizes | Data Design | This study reports student achievement data as: | Data_Design | 1) Pre-post w/control  
2) Pre-post w/o control  
3) Gain score w/control  
4) Gain score w/o control  
5) Post-only w/control  
6) Correlations (Regression coefficients, HLM, etc)  
7) Percent passing/proficient  
8) combination of above | To find the rating, add up the number of "yes" responses under the Study Quality Contract. Should range from 0 to 6. |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality</td>
<td>What is the quality rating of this study?</td>
<td>Quality</td>
<td>Numerical</td>
<td>N/A</td>
</tr>
<tr>
<td>Number of Subgroups</td>
<td>How many subgroups is there achievement data for?</td>
<td>N_Subgroups</td>
<td>Numerical</td>
<td>Subgroups may include: Gender, ethnicity, SES, Special</td>
</tr>
<tr>
<td>Number of Measures</td>
<td>How many student achievement measures were used in the study?</td>
<td>N_Measures</td>
<td>Numerical</td>
<td>Measures may include: standardized test, research-developed test, content sub-tests, etc.</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------------------------------------------------------</td>
<td>------------</td>
<td>----------</td>
<td>------------------------------------------------------------------</td>
</tr>
<tr>
<td>Number of Effects</td>
<td>How many effects of PD on student achievement were reported?</td>
<td>N_Effects</td>
<td>Numerical</td>
<td>To find this: Multiply N_Subgroups by N_Measures.</td>
</tr>
<tr>
<td>PD Characteristics</td>
<td>Intervention Name</td>
<td>Specify the name (if any) of the intervention</td>
<td>PD_Name</td>
<td>String</td>
</tr>
<tr>
<td></td>
<td>Content Focus</td>
<td>Specify whether or not increasing content knowledge was a specific focus of the PD</td>
<td>Focus_Content</td>
<td>1) Yes 2) No</td>
</tr>
<tr>
<td>Focus Area</td>
<td>Question</td>
<td>Focus Code</td>
<td>Options</td>
<td>Notes</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>------------------</td>
<td>---------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Shared Vision Focus</td>
<td>Specify whether or not creating a shared vision was a specific focus of the PD</td>
<td>Focus_Vision</td>
<td>1) Yes 2) No</td>
<td>This includes focii determined by the coder and listed by the author.</td>
</tr>
<tr>
<td>Student Learning Focus</td>
<td>Specify whether or not learning how students learn was a specific focus of the PD</td>
<td>Focus_Learn</td>
<td>1) Yes 2) No</td>
<td>This includes focii determined by the coder and listed by the author.</td>
</tr>
<tr>
<td>Pedagogical Content Knowledge Focus</td>
<td>Specify whether or not increasing pedagogical content knowledge was a specific focus of the PD</td>
<td>Focus_PCK</td>
<td>1) Yes 2) No</td>
<td>This includes focii determined by the coder and listed by the author.</td>
</tr>
<tr>
<td>Sense of Self as Math Teacher Focus</td>
<td>Specify whether or not increasing the teachers' sense of self as a math teacher was a specific focus of the PD</td>
<td>Focus_Self</td>
<td>1) Yes 2) No</td>
<td>This includes focii determined by the coder and listed by the author.</td>
</tr>
<tr>
<td>Equity Focus</td>
<td>Specify whether or not increasing equity knowledge was a specific focus of the PD</td>
<td>Focus_Equity</td>
<td>1) Yes 2) No</td>
<td>This includes focii determined by the coder and listed by the author.</td>
</tr>
<tr>
<td>PD Characteristics</td>
<td>Specify whether or not increasing knowledge of cooperative learning focus of the PD</td>
<td>Focus_CoLearn</td>
<td>1) Yes</td>
<td>2) No</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>---------------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>Cooperative Learning Focus</td>
<td>Specify whether or not increasing classroom management knowledge was a specific focus of the PD</td>
<td>Focus_Manage</td>
<td>1) Yes</td>
<td>2) No</td>
</tr>
<tr>
<td>Classroom Management Focus</td>
<td>Specify any other focuses of PD not listed above.</td>
<td>PD_Focus_Other</td>
<td>String</td>
<td></td>
</tr>
<tr>
<td>PD Characteristics Other</td>
<td>Specify whether or not the PD was designed specifically to target math achievement</td>
<td>Math_PD</td>
<td>1) Yes</td>
<td>2) No</td>
</tr>
<tr>
<td>Math Specific PD</td>
<td>This includes foci determined by the coder and listed by the author.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>This includes foci determined by the coder and listed by the author.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>If a focus of the PD cannot fit into one of the other categories, please list that focus here. (eg: classroom management)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

107
<table>
<thead>
<tr>
<th>Math Content Area</th>
<th>Specify which math sub-content areas were targeted</th>
<th>Math_Content</th>
<th>1) counting 2) computations 3) number facts 4) algebra 5) geometry 6) word problems/problem solving 7) other</th>
<th>Choose only one answer. If there is more than one focus, list them under &quot;other&quot;.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other Math Content</td>
<td>Describe the &quot;other&quot; math content area focus</td>
<td>Content_Other</td>
<td>String</td>
<td>Only respond to this if Math_Content = 7</td>
</tr>
<tr>
<td>Duration and Frequency</td>
<td>Describe in detail the duration and frequency of the PD</td>
<td>PD_Time</td>
<td>String</td>
<td>For example: two full days of PD followed by weekly hour-long team meetings</td>
</tr>
<tr>
<td>Fidelity of Implementation</td>
<td>Was Fidelity of Implementation Described in the Study</td>
<td>Fidelity</td>
<td>1) Yes 2) No</td>
<td></td>
</tr>
<tr>
<td>Fidelity of Implementation Description</td>
<td>Describe fidelity of implementation</td>
<td>Fidelity_Desc</td>
<td>String</td>
<td></td>
</tr>
</tbody>
</table>
| PD Development | Who developed the intervention? | PD_Develop | 1) designed by the authors  
2) commercially developed  
3) developed at a university  
4) other |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PD Development Other</td>
<td>Specify the &quot;other&quot; who developed the intervention</td>
<td>Develop_Other</td>
<td>String</td>
</tr>
</tbody>
</table>
| Voluntary Participation | Was the participation in the PD intervention Voluntary? | PD_Voluntary | 1) Yes  
2) No  
3) Not indicated |
| PD Delivery | Who delivered the PD? | PD_Delivery | 1) the researcher(s)  
2) the developers (if different from the researchers)  
3) school administration  
4) other teachers ("lead teacher" format)  
5) not indicated  
6) other |
| PD Delivery Other | What "other" delivered the PD? | Delivery_Other | String |

Only respond to this if PD_Develop = 4

Only respond to this if PD_Delivery = 6
<table>
<thead>
<tr>
<th><strong>Student Characteristics</strong></th>
<th><strong>Treatment Student Grade</strong></th>
<th>Which Grade(s) did the treatment students come from?</th>
<th>T_SS_Grade</th>
<th>String</th>
<th>List all that apply or put &quot;not indicated&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Student Grade</td>
<td></td>
<td>Which Grade(s) did the Control students come from?</td>
<td>C_SS_Grade</td>
<td>String</td>
<td>List all that apply or put &quot;not indicated&quot;</td>
</tr>
<tr>
<td>Treatment Age Mean</td>
<td></td>
<td>What was the mean age of the treatment students?</td>
<td>T_Age_Mean</td>
<td>Numerical</td>
<td></td>
</tr>
<tr>
<td>Control Age Mean</td>
<td></td>
<td>What was the mean age of the control students?</td>
<td>C_Age_Mean</td>
<td>Numerical</td>
<td></td>
</tr>
<tr>
<td>Treatment Age SD</td>
<td></td>
<td>What was the standard deviation of ages of the treatment students?</td>
<td>T_Age_SD</td>
<td>Numerical</td>
<td></td>
</tr>
<tr>
<td>Control Age SD</td>
<td></td>
<td>What was the standard deviation of ages of the control students?</td>
<td>C_Age_SD</td>
<td>Numerical</td>
<td></td>
</tr>
<tr>
<td>Treatment Student Female</td>
<td></td>
<td>What percentage of the treatment students were female?</td>
<td>T_SS_Female</td>
<td>Numerical</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Question</td>
<td>Code</td>
<td>Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>---------</td>
<td>----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Student Female</td>
<td>What percentage of the control students were female?</td>
<td>C_SS_Female</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment Students Male</td>
<td>What percentage of the treatment students were male?</td>
<td>T_SS_Male</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Students Male</td>
<td>What percentage of the control students were male?</td>
<td>C_SS_Male</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment Students African American</td>
<td>What percentage of the treatment students were African American?</td>
<td>T_SS_Black</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Students African American</td>
<td>What percentage of the control students were African American?</td>
<td>C_SS_Black</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment Students Asian American?</td>
<td>What percentage of the treatment students were Asian American?</td>
<td>T_SS_Asian</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Students Asian American</td>
<td>What percentage of the control students were Asian American?</td>
<td>C_SS_Asian</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Question</td>
<td>Code</td>
<td>Type</td>
<td>Constraints</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>----------------------------------------------------</td>
<td>--------</td>
<td>----------</td>
<td>-----------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Treatment Students White</td>
<td>What percentage of the treatment students were White?</td>
<td>T_SS_White</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Students White</td>
<td>What percentage of the control students were White?</td>
<td>C_SS_White</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment Students Hispanic</td>
<td>What percentage of the treatment students were Hispanic?</td>
<td>T_SS_Hisp</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Students Hispanic</td>
<td>What percentage of the control students were Hispanic?</td>
<td>C_SS_Hisp</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SES Indicated</td>
<td>Is SES indicated?</td>
<td>SS SES</td>
<td>1) Yes, 2) No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SES Described</td>
<td>Describe the categorization of SES and corresponding descriptive statistics</td>
<td>SS SES_Desc</td>
<td>String</td>
<td>Only respond if SS SES = 1</td>
<td></td>
</tr>
<tr>
<td>Teacher Characteristics</td>
<td>Treatment Teachers Female What percentage of the treatment Teachers were female?</td>
<td>T_Tch_Female</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>Question</td>
<td>Code</td>
<td>Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-----------</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Teachers Female</td>
<td>What percentage of the control Teachers were female?</td>
<td>C_Tch_Female</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment Teachers Male</td>
<td>What percentage of the treatment Teachers were male?</td>
<td>T_Tch_Male</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Teachers Male</td>
<td>What percentage of the control Teachers were male?</td>
<td>C_Tch_Male</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment Teachers African American</td>
<td>What percentage of the treatment Teachers were African American?</td>
<td>T_Tch_Black</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Teachers African American</td>
<td>What percentage of the control Teachers were African American?</td>
<td>C_Tch_Black</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment Teachers Asian American?</td>
<td>What percentage of the treatment Teachers were Asian American?</td>
<td>T_Tch_Asian</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Teachers Asian American</td>
<td>What percentage of the control Teachers were Asian American?</td>
<td>C_Tch_Asian</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Treatment Teachers White</strong></td>
<td>What percentage of the treatment Teachers were White?</td>
<td>T_Tch_White</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------------------------------------------------</td>
<td>-------------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Control Teachers White</strong></td>
<td>What percentage of the control Teachers were White?</td>
<td>C_Tch_White</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Treatment Teachers Hispanic</strong></td>
<td>What percentage of the treatment Teachers were Hispanic?</td>
<td>T_Tch_Hisp</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Control Teachers Hispanic</strong></td>
<td>What percentage of the control Teachers were Hispanic?</td>
<td>C_Tch_Hisp</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SES Indicated</strong></td>
<td>Is SES indicated?</td>
<td>Tch SES</td>
<td>1) Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2) No</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SES Described</strong></td>
<td>Describe the categorization of SES and corresponding descriptive statistics</td>
<td>Tch_SES_Desc</td>
<td>String</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Only respond if SS_SES = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Treatment Teachers High School</strong></td>
<td>What percentage of the treatment Teachers had at most a high school degree?</td>
<td>T_Tch_HS</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>Question</td>
<td>Code</td>
<td>Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>------</td>
<td>------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Teachers High School</td>
<td>What percentage of the control Teachers had at most a high school degree?</td>
<td>C_Tch_HS</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment Teachers Some College</td>
<td>What percentage of the treatment Teachers had at most some college?</td>
<td>T_Tch_Scollege</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Teachers Some College</td>
<td>What percentage of the control Teachers had at most some college?</td>
<td>C_Tch_Scollege</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment Teachers College</td>
<td>What percentage of the treatment Teachers had at most a college degree?</td>
<td>T_Tch_College</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Teachers College</td>
<td>What percentage of the control Teachers had at most a college degree?</td>
<td>C_Tch_College</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment Teacher Master's</td>
<td>What percentage of the treatment Teachers had at most a Master's Degree?</td>
<td>T_Tch_Masters</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----------------------------------------------------------------</td>
<td>------------------------------</td>
<td>------------------------------</td>
<td>------------------------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>Control Teachers Master's</td>
<td>What percentage of the control Teachers had at most a Master's</td>
<td>C_Tch_Masters</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Degree?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment Teachers PhD</td>
<td>What percentage of the treatment Teachers had a doctoral degree?</td>
<td>T_Tch_PhD</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Teachers PhD</td>
<td>What percentage of the control Teachers had a doctoral degree?</td>
<td>C_Tch_PhD</td>
<td>Numerical</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Are the teachers' years of experience indicated?</td>
<td>Exp_Indicated</td>
<td>1) Yes 2) No</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Describe the categorization and descriptive statistics for the</td>
<td>Exp_Desc</td>
<td>String</td>
<td>Only answer this if Exp_Indicated = 1</td>
<td>This is often indicated with a min, max, Mean, and SD.</td>
</tr>
<tr>
<td></td>
<td>years of experience</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous PD</td>
<td>Is previous PD experience indicated?</td>
<td>PrePD_Indicated</td>
<td>1) Yes 2) No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous PD Described</td>
<td>Describe the categorization and descriptive statistics for previous PD experience</td>
<td>PrePD_Desc</td>
<td>String</td>
<td>Only answer this if PrePD_Indicated = 1</td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>------------</td>
<td>--------</td>
<td>----------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Teacher Test</td>
<td>Were teachers administered a test of knowledge?</td>
<td>Tch_Test</td>
<td>1) Yes  2) No</td>
<td>Only answer this if Tch_Test = 1</td>
<td></td>
</tr>
<tr>
<td>Teacher Test Described</td>
<td>Describe the teacher test and any descriptive results</td>
<td>Tch_Test_Desc</td>
<td>String</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher Certification</td>
<td>Is teacher certification indicated?</td>
<td>Tch_Cert</td>
<td>1) Yes  2) No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher Certification Described</td>
<td>Describe the teacher certification categorization and any descriptive statistics</td>
<td>Tch_Cert_Desc</td>
<td>String</td>
<td>Only answer this if Tch_Cert_Desc = 1</td>
<td></td>
</tr>
</tbody>
</table>
### Appendix B: Effect-size Data Table

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Subgroup</th>
<th>Raw or Gain?</th>
<th>Comparison Group</th>
<th>Control</th>
<th>Treatment 1</th>
<th>Treatment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Describe Treatment</th>
<th>Pre</th>
<th>Post</th>
<th>Pre</th>
<th>Post</th>
<th>Pre</th>
<th>Post</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t-value</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>F-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Page #             |     |      |     |      |     |      |