7-1-2022

The Efficacy of Research-Based "mathematics for All" Professional

Kathleen Melhuish
Texas State University

Eva P. Thanheiser
Portland State University, evat@pdx.edu

Alexander White
Texas State University

Brenda Rosencrans
Portland State University

J. Michael Shaughnessy
Teacher Development Group

See next page for additional authors

Citation Details

Melhuish, Kathleen; Thanheiser, Eva P.; White, Alexander; Rosencrans, Brenda; Shaughnessy, J. Michael; Foreman, Linda; Riffel, Andrew; and Guyot, Layla, "The Efficacy of Research-Based "mathematics for All" Professional" (2022). *Mathematics and Statistics Faculty Publications and Presentations*. 354.
https://pdxscholar.library.pdx.edu/mth_fac/354

Authors
Kathleen Melhuish, Eva P. Thanheiser, Alexander White, Brenda Rosencrans, J. Michael Shaughnessy, Linda Foreman, Andrew Riffel, and Layla Guyot

The Efficacy of Research-Based “Mathematics for All” Professional Development

Kathleen Melhuish

Texas State University

Eva Thanheiser

Portland State University

Alex White

Texas State University

Brenda Rosencrans

Portland State University

J. Michael Shaughnessy

Teacher Development Group

Linda Foreman

Teacher Development Group

Andrew Riffel

Portland State University

Layla Guyot

Texas State University

Author Note

Corresponding Author: Kathleen Melhuish, melhuish@txstate.edu, Mathematics Department, Texas State University, 601 University Dr., San Marcos, TX.

This material is based upon work supported by the National Science Foundation under Grant No. DRL-1223074. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Abstract

This paper contributes to the larger narrative around what makes a mathematics professional development (PD) successful and in what ways. We share a research-based PD model that was implemented in elementary schools in an urban school district for three years. The model uses a pseudo-lesson study approach and emphasizes standards-based instruction. We found that teachers made gains in knowledge and instruction quality. However, while some students saw gains on standardized assessments, this was only the case for students who were not members of historically minoritized groups (Black/Latino) countering our assumptions that the PD would lead to equitable achievement results. We conclude with a discussion of how a color-blind approach to PD may account for the inequitable results.

Keywords: Professional Development, Lesson Study, Achievement Equity

1 Introduction

As a field, we have reached a general consensus on the features of professional development (PD) programs that can engender teacher change and impact student learning (e.g., Darling-Hammond & Richardson, 2009; Desimone, 2009). However, studying these programs at-scale is challenging, and there remains much to learn about how well-designed PDs ultimately affect both teachers and students (e.g, Sztajn et al., 2017). In the context of mathematics, many PD programs take a standards-based approach to instruction (National Council of Teachers of Mathematics [NCTM], 2000, 2014) emphasizing student thinking, conceptual understanding, and promoting student engagement in mathematical practices (Goldsmith et al., 2013). Within our community, there has been extensive debate (e.g., Barajas-López, & Larnell, 2019; Martin, 2007; Rubel, 2017) as to whether such approaches are sufficient for more equitable classrooms. If we consider student outcomes (e.g., achievement dimension of equity; Gutiérrez, 2012), several well-known studies have pointed to standards-based instruction leading to equitable outcomes for students of varying racial and ethnic backgrounds (e.g., Boaler & Staples, 2008; Silver, & Stein, 1996). In contrast, a number of qualitative studies have pointed to inequitable experiences for students with different racial, socioeconomic, or gender backgrounds (e.g., Esmonde et al., 2009; Lubienski, 2002; Murrell, 1994).

In this paper, we share a large-scale study of a complex-instruction informed (Cohen et al., 1999), standards-based PD designed to align with research on successful PD programs (e.g., Darling-Hammond & Richardson, 2009; Desimone, 2009). The PD emphasizes a non-deficit view of students, a status-free mathematics classroom, high cognitive demand tasks where students justify and generalize, and teaching routines and actions to engage students in mathematical practices and rich discussion. The PD was implemented for three years with grades

3-5 teachers in an urban school district. The teachers participated in either a full (pseudo-lesson studies and summer workshop: FullPD) or a summer workshops only (SummerPD) model.

Based on prior studies, we anticipated the FullPD would lead to (1) teacher knowledge gains, (2) gains in quality of instruction, and (3) student achievement gains. Further, we entered the project with a " more than one road to equity" (Rubel, 2017, p. 77) approach anticipating a (4) replication of the more equitable outcomes found in Boaler and Staples' (2008) study which informed the instruction emphasized in the PD. By focusing on all four change goals, we are able to provide a more nuanced analysis that goes beyond the common teacher-only focus of many PD studies (Sztajn et al., 2017).

We explore the following research questions:

1. Do (and to what degree do) FullPD and SummerPD teachers make gains in their mathematical knowledge for teaching (MKT) (Hill et al., 2004)?
2. Do (and to what degree do) FullPD and SummerPD teachers make gains in their quality of mathematical instruction (MQI) (Hill, 2014)?
3. How are student standardized assessment scores related to their time in FullPD and SummerPD classrooms?
4. How are standardized assessment scores related to student backgrounds, and to what degree does time in FullPD or SummerPD classrooms lead to more equitable scores?

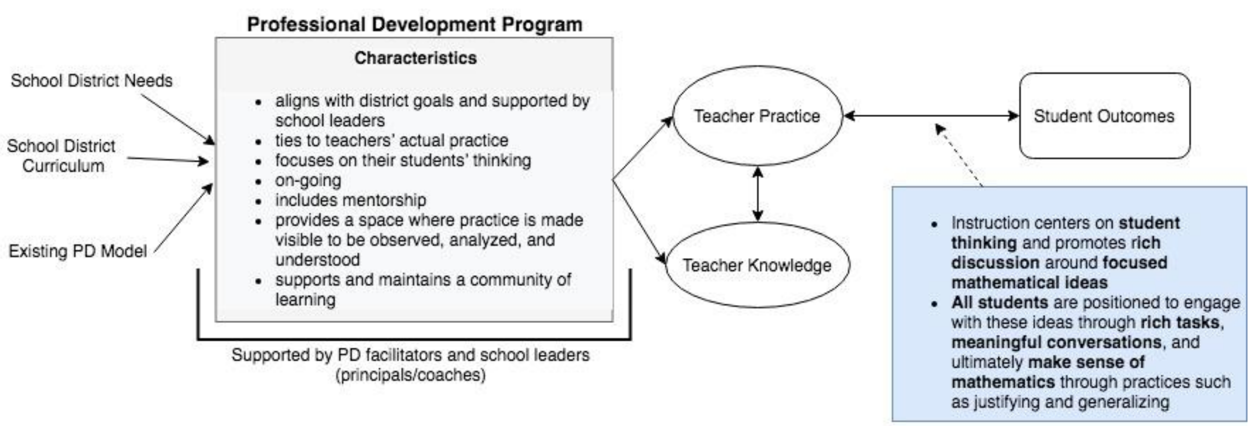
We briefly share a series of results that align with our research-based expectations: teachers demonstrated mathematical knowledge for teaching (MKT) gains in both PD models and only FullPD teachers had gains in their mathematical quality of instruction (MQI). We then consider results from a more complex analysis of student assessment scores. These results reflect several expected outcomes (e.g., implementing more standards-based instruction was linked to higher estimated assessment scores), but also an unexpected outcome: a widening opportunity gap for minoritized students. We structure the paper first by introducing the literature and theory that informed our PD and initial hypotheses. We then share the results of this study. We conclude with a return to the literature to make sense of the ways that re-producing the common supposedly “race-neutral” approach to equity often found in mathematics education (Barajas-López & Larnell, 2019) may account for an inequitable result of an otherwise successful PD.

2 Theoretical Assumptions and Background

In this section, we outline assumptions about PD and teacher change, and relevant literature, that influenced the design and enactment of the PD model. The full framing can be found in Figure 1.

Figure 1

Components of our conceptual framework



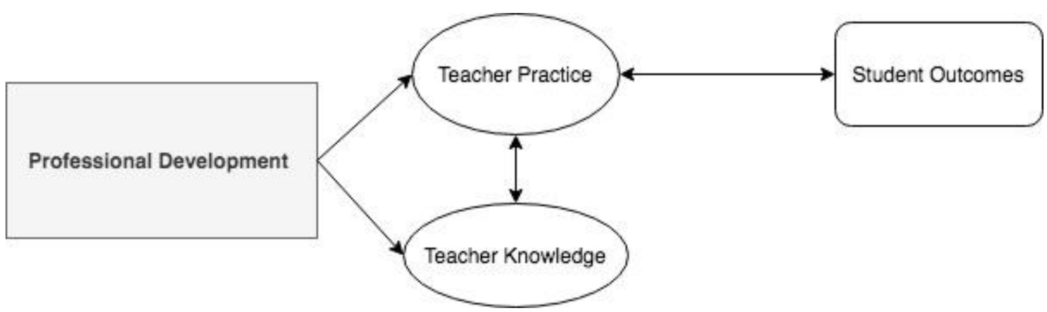
2.1 Assumptions about learning and the classroom

A social constructivist stance underlies our project: students build knowledge and construct meaning while engaging with their peers and teachers. This stance aligns with many of the components of previously successful, standards-based PD programs (Goldsmith et al., 2013). Our assumption is that student thinking and voices should drive mathematics**.** By centering student thinking, their sense-making becomes the focus of instruction supporting ownership of the mathematics and conceptual understanding. Further, rich and focused tasks are essential to provide opportunities for students to engage in mathematical thinking and practices. Teachers then serve a fundamental role in shaping student opportunities in the classroom both in terms of task enactment and through the development of norms where students engage with each other’s ideas (e.g., Staples, 2007). This stance on learning and the classroom aligns with numerous PD efforts including those focused on teachers’ noticing of student thinking (e.g., Jacobs et al., 2010; Sherin & van Es, 2005), selecting high demand tasks and maintaining high demand during classroom implementation (e.g., Boston & Smith, 2009), and promoting productive teacher moves can be powerful for student learning (e.g., Michaels & O'Connor, 2015).

2.2 Assumptions about teacher change

Figure 2

Relationship between PD, teacher practice, teacher knowledge, and student outcomes



Taking a broad stance, we see PD as a potential impetus for teacher change where change is operationalized as growth (Clarke & Hollingsworth, 2002). In this work, we focus on growth in knowledge and practice. While some models of teacher change focus on a linear process where teacher practice is changed via a PD directly, we make the assumption that a PD can lead to change both directly and indirectly (Clarke & Hollingsworth, 2002). A PD may lead to a teacher making a small change in practice. Through making this change, the teacher may see an impact on their students which engenders a desire to maintain changes and make additional change in practice. Further, the focus of a PD may directly impact a teacher's pedagogical knowledge, but changes in practice could also lead to experiences that impact knowledge (See Figure 2). We also acknowledge that this change occurs in conjunction with the surrounding context. PD is unlikely to lead to changes without a supportive surrounding community and content that is relevant to a teacher’s practice and setting (see Figure 1). In the next sections, we unpack these assumptions.

2.2.1 Teacher change is dependent on seeing change as relevant and connected to their classrooms. In general, teachers learn when motivated to do so, and when they engage with authentic situations connected to their teaching (Ball & Cohen, 1999). Teachers are more likely to *implement* what they are learning when PD *models* what they are learning, especially when it is situated in their own classrooms (Snow-Renner & Lauer, 2005) or brings in and builds on artifacts of children’s thinking (Carpenter et al., 1989; Desimone et al., 2002; Penuel et al., 2007). In addition, teachers strongly associate hands-on, active work with the value of the PD (Darling-Hammond et al., 2009).

2.2.2 Teacher change occurs within the larger system of school and district priorities and initiatives. PDs are more successful when they align with local priorities and initiatives (Darling-Hammond et al., 2009; Desimone et al., 2002; Garet et al., 2001; Penuel et al., 2007; Snow-Renner & Lauer, 2005). Coherence across PD programs and alignment with district goals, standards, and other influences reduces the push and pull effects on teachers as to which direction to go (Snow-Renner & Lauer, 2005). Coherence to district priorities and initiatives provide teachers with a way to incorporate what they are learning in the PD into their practice and larger school/district context without feeling like this is yet “another thing” that they need to do (Garet et al., 2001).

2.2.4 Teacher change can be supported and maintained through a community of learning amongst teachers and leaders at the school. To provide teachers with settings and resources needed to engender and maintain change, it is essential to develop collaborative relationships amongst teachers. Stoll et al. (2006) note that we have moved away from changing individual teachers to models where “whole school communities need to work and learn together to take charge of change” (p. 222). *Professional learning communities (PLC)* are a promising route for developing the collaboration needed for systematic and sustained change. In this context, a PLC is a group of teachers (and other school members) who collaboratively engage in critical examination of their practice in order to learn and grow (e.g., Toole & Louis, 2002). A major component of PLC is then deprivatizing practice (Louis et al., 1996) where practice is repositioned as communal instead of individual, allowing for joint planning, observing, and reflecting (Darling-Hammond et al., 2009; Fullan, 2007).

The lesson study approach, which underlies the FullPD, provides one means to de-privatize practice. In this model, teachers jointly plan, observe implementation, and debrief lessons. Particularly high impact factors of lesson study include: (1) identifying a mathematical goal for a lesson, (2) collective planning, (3) collectively observing a live lesson, and (4) collective reflection and refinement (Fernandez, 2002). Although lesson study was developed in the context of the Japanese education system, researchers have shown that this model can similarly impact teachers’ and students’ knowledge in the United States (Lewis & Perry, 2017).

2.2.5. Change takes time and continued support. Rather than one-off workshops which focus on various ideas, PD programs should be ongoing and long-term (Garet et al., 1999) so that teachers are able to learn about a practice, attempt implementation, and revisit the practice multiple times if needed (Darling-Hammond et al., 2009). Teachers benefit more from PD “when their learning is reinforced over time through repeated and varied exposure to ideas and through interactions with colleagues, who can act as a resource for each other’s learning.” (Knapp, 2003, p. 121). In addition, more intense and sustained PDs have resulted in greater student achievement gains (Corcoran et al., 2003; Supovitz & Turner, 2000).

2.2.6 Coaching and mentorship provide a mechanism to support teachers’ continued learning. Coaching is often recommended as part of a PD program where coaches collaborate with teachers and model lesson planning (Kennedy, 2016) and implementation. Coaching has the potential to affect the ability of teachers to analyze their own instruction and has been documented to increase teachers’ MQI (Kraft & Hill, 2018) and impact student outcomes (Campbell & Malkus, 2011; Neufeld & Roper, 2003).

2.2.7 For teacher change to be connected to their students, PD must provide opportunity for teachers to actively engage with content and student thinking. Professional development programs are more effective when they focus on specific subject (e.g., mathematics) content (Blank et al., 2007; Garet et al., 2001; Marek & Methven, 1991) and specify the concepts and skills to be learned by students and the likely difficulties students will encounter with those concepts and skills (Ball & Cohen, 1999; Cohen & Hill, 2001; Darling-Hammond et al., 2009). Focusing on enhancing teachers' understanding of content, student thinking, and student motivation can lead to increased student conceptual understanding (Carpenter et al., 1989; Saxe et al., 2001).

2.3 Hypothesized Outcomes

As a result of the prior empirical work that informed the design of our study, we entered the project with several key hypotheses. First, we hypothesized that *teachers would increase their mathematical knowledge for teaching* as a result of engaging in the PD (as measured by the Learning Mathematics for Teaching (LMT) project’s (Hill et al., 2005) MKT assessment focused on the knowledge needed to teach number and operation). We selected this instrument because of the validation process (Schilling et al., 2007), its strong development, its common use in the field, and its alignment with knowledge the project valued. Further, MKT has been linked to student outcomes and quality of instruction (Hill, et al., 2012). Many prior PD studies that focus on student thinking have demonstrated gains using instruments from the LMT project (e.g., Bell et al., 2010; Moyer-Packenham & Westenskow, 2012).

Second, we hypothesized that *teachers would increase their instructional quality* as measured by the Mathematical Quality of Instruction (MQI) instrument (Hill, 2014). We selected the MQI both because of its extensive validation work and prominence in the field and also its alignment with several important aspects of standards-based instruction and our PD model. The instrument has dimensions: richness of mathematics, errors and imprecision, working with students and mathematics, and common core aligned student practices. PDs with similar focus have documented gains in participating teachers’ MQI (Koellner & Jacobs, 2015); however, other studies have pointed to the fact that MKT gains, and a standards-based PD, may be insufficient to also see MQI gains (Jacob, et al., 2017).

From the student side, we hypothesized that the changes in instruction would *lead to students gaining a stronger understanding of the mathematics* in these classrooms (as reflected on standardized assessments). PD and instructional intervention projects with similar emphases have documented such gains (Boaler & Staples, 2008; Hill et al., 2005; Silver & Stein, 1996).

Further, we hypothesized that these gains would occur for *all groups of students and the type of instruction emphasized would lead to more equitable outcomes*. By focusing on students' mathematical thinking, positioning students as contributors to mathematics, and centering peer-to-peer argumentation, we replicated common ties to characterizing more equitable classrooms (Jackson & Cobb, 2010). Similar interventions with a focus on students engaging actively in rich mathematics (e.g., the Railside School study, Boaler & Staples, 2008); the QUASAR study, Silver & Stein, 1996) have pointed to the potential for these types of interventions to lead to gains for all students. Boaler and Staples found that differences in assessment outcomes between white students and students of color disappeared after several years in a school emphasizing justification and multiple solution strategies. Even on standardized assessments, the differences between groups shrank. These teachers were also trained in complex instruction where students took on group roles and teachers assigned competency. Silver and Stein also documented this trend, even without inclusion of complex instruction, in their work with urban middle school teachers emphasizing high cognitive demand tasks with a focus on multiple solutions, multiple representations, and collaboration. Students in these schools made substantial gains on their assessments with white and Black students making equal gains in some cohorts and Black students making greater gains in others. Studies of standards-based curricula studies point to similar results with all students making gains, but Black and Latino/Hispanic[1] students making larger gains (e.g., Riordan & Noyce, 2001). A focus on understanding, especially when paired for a focus on ideas over language, has also been advocated for transitional bilinguals[2]

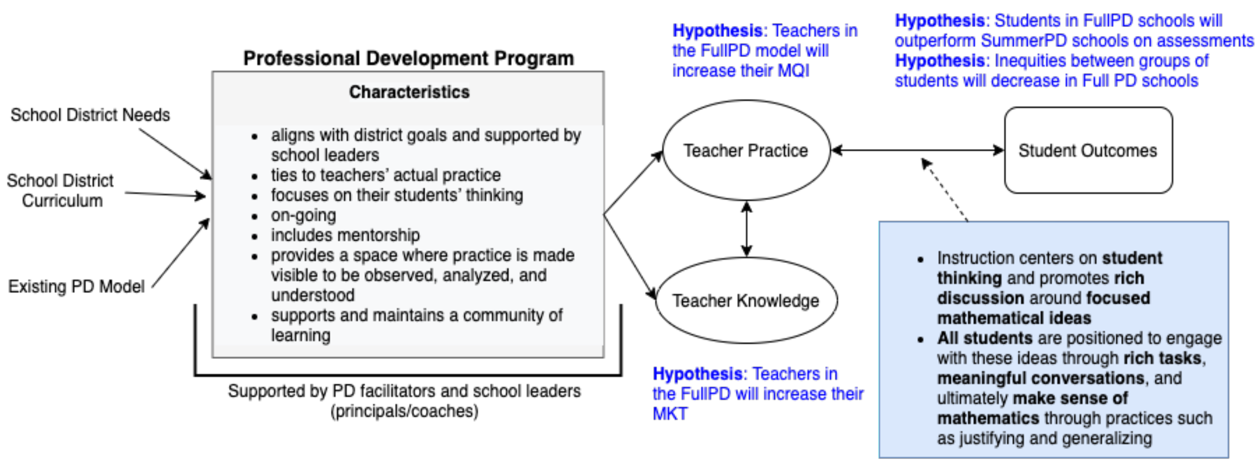
1

2 We use Black, Latino/Hispanic, and transitional bilingual throughout the manuscript in alignment with the participating school district's language.

(Moschkovich, 2013) and evidenced to lead to learning gains amongst this population (Staples & Truxaw, 2010). In general, these studies point to the potential of standards-based, conceptually-focused PDs in leading to more equitable achievement outcomes. See Figure 3 for hypotheses.

Figure 3

Key hypotheses



3 Methods and Context

In this section, we provide both the details of our PD model and the methods and context for our efficacy study.

3.1 The Professional Development Structure and Context

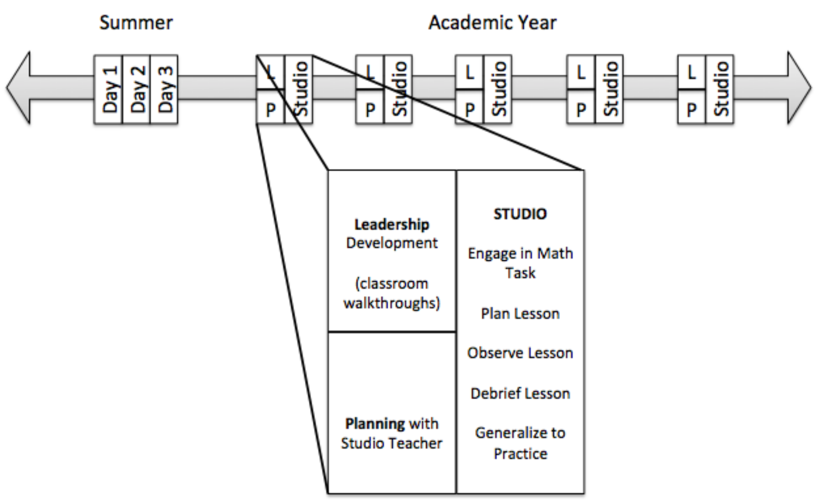
The PD took place in an urban school district in the United States. The PD developers worked with the district leaders and principals to connect and create coherence amongst the district's initiatives including working with local instructional coaches (one at each school), and supporting coherence with the existing (non-math specific) professional learning communities in the schools. Further, the PD itself was integrated into tasks and lessons existing within the school districts' curriculum.

3.1.1 The Structure and Content of the PD. The PD model is characterized by two types of PD sessions: Best Practices (BP) summer workshops, and year-round lesson cycles. This project took place over 3 years. The content of the PD (Melhuish & Thanheiser, 2017) focused on the promotion of a set of standards-based *teaching routines* (working with public records of students mathematical thinking, selecting and sequencing student mathematical ideas, working with misconceptions and errors, working with visual representations, structuring student talk, and conferring with students to understand their thinking) and a set of core *habits of mind and interaction that* operationalize ways for students to engage with each other and the mathematics in alignment with mathematical practices (e.g., reasoning with representations, justifying, generalizing, critiquing and debating.)

During each summer, teachers, coaches, and principals (in both the FullPD and SummerPD model) attended 3-day workshops focused on best practices in standards-based instruction. Year 1 focused on the focal goals of engaging students in justifying and generalizing, describing the other habits of mind and interaction, introducing teaching routines, and planning high cognitive demand tasks. Year 2's workshop focused on the development of a culture where students engage productively, including a continued emphasis on all students being capable of engaging in the habits of mind and interaction. The materials also focused on equity and the teacher's role in creating status-free classrooms. All teachers read and debriefed an article about mindset (Dweck, 2007) and an article that focused on promoting status-free equitable classrooms via complex instruction (Boaler, 2006). The focus of Year 3's workshop was whole number operations and how teachers (through teaching routines) can support students in engaging deeply with whole number operations via the habits of mind and interaction. During these sessions, teachers watched videos of students engaging in mathematics, analyzed student work, engaged in mathematical tasks themselves (where PD facilitators modeled and identified important instructional practices), read relevant articles, and worked together to debrief and think about how this work could impact their teaching practice.

Figure 4

Structure of one year of PD for FullPD.



The FullPD included both the summer workshops and studios that occur across five 2-day sessions during the academic year (See Figure 4). At each school, one teacher served as the *studio teacher* opening their classroom to their fellow teachers, the *resident teachers*. On the morning of the first day, a facilitator worked with the principal (and school coach) developing their leadership voice and school organization. The facilitator, principal, and coaches also observed classrooms together in order to discuss formative evaluation based on the degree teachers were implementing teaching routines and supporting students in habits of mind/interaction. During the second half of the first day, the facilitator and studio teacher spent time lesson planning. On the second day, all participants attended. The resident teachers refined and finalized the lesson plan by doing the mathematics and anticipating student thinking. Then resident teachers observed the studio teacher implement the lesson. After the lesson, all participating teachers debriefed the lesson. The debrief often focused on data collected about the students’ mathematical reasoning and what type of teacher moves may have supported the students’ reasoning (see Melhuish, et al., 2020). They concluded by reflecting on what they learned and how to connect that to their own practice. In contrast to traditional lesson study, the lessons’ foci were situated within the teachers’ current topic in the curriculum. In this way, the model is adaptive to where teachers are at, but does not involve re-implementing the same lesson. This model incorporates high leverage practices of lesson study (Fernandez, 2002), without the necessity of producing a finalized shared lesson -- which may not always be feasible in the confines of traditional US school districts.

3.2 The Study Design

In this section, we outline the design of the study including demographics of the participating schools and information about data collection.

3.2.1 School Demographics. We used a stratified, random cluster approach to assign an initial set of 27 elementary schools (that had opted into or out of the study based on principal and teacher interest) to either the FullPD or SummerPD models based on percentage of students eligible for free and reduced lunch (FRL) and percentage of students meeting standards on the prior years’ mathematics standardized assessments (MSP). The strata reflected low (below 50%), medium (50-75%), and high (above 75%) levels of FRL and MSP. We then randomly assigned treatments within strata with the exception of 3 schools not assigned to FullPD because of low consent. These numbers changed during the implementation of the PD with two FullPD schools withdrawing and the remaining schools in the district opting into the SummerPD[3]. We share data from a total of 33 elementary schools: 20 SummerPD, and 13 FullPD. Information on these schools can be found in Tables 1 and 2. Even with fluctuations in participating schools, there was no significant difference between demographic categories across SummerPD and FullPD schools.

Table 1

School Demographics Related to Assignment at Start of Project Disaggregated by Treatment

		Percentage Free or Reduced Lunch Eligible		Percentage Meeting Standards on Math Assessment	
	n	*Mean*	*SD*	*Mean*	*SD*
Total	33	67.5	23.6	55.2	16.5
Summer PD	20	69.1	22.0	53.7	16.7
FullPD	13	65.0	26.7	56.9	17.3

Table 2

School Demographics Related to Race/Ethnicity at Start of Project

		Percentage of Students Black/African American		Percentage of Students Hispanic/Latino		Percentage of Students Asian	
	n	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Total	33	22.1	10.4	15.6	9.6	10.9	5.4
Summer PD	20	22.0	11.6	15.9	10.1	12.2	5.8
FullPD	13	22.3	8.7	15.3	9.1	8.9	4.2

[3] For practical reasons, the process was not truly random. Additional analysis verifying similarity between FullPD and SummerPD schools can be found in the supplemental material.

3.2.2 Data Collection Related to Measures. We measured participating teachers’ MKT during the first day of the three summer workshops. These occurred during summer before the PD (baseline), then the following two summers (after year 1 and 2 of the project). We used the LMT’s multiple choice assessment focused on number and operation which includes two score equivalent versions. Teachers did not take the same version in consecutive years. The MKT scores are assigned based on an item response theory model (person-item reliability of 0.91, Hill, 2010) and are interpretable as standard deviations.

All teachers were asked to participate in videotaping classrooms, with 94 agreeing. For each of these teachers, we videotaped two lessons at the end of the year using the protocol from the Mathematical Quality of Instruction (MQI) (Hill, 2014) developers. These lessons occurred at the end of baseline year, year 1, year 2, and year 3 of the project. For this paper, we focus on the more complete data set from earlier in the project. Videos were coded by a team of coders who certified through Harvard University’s certification process. While coding occurred, we calibrated through weekly seminars where all coders coded the same two five-minute segments. Any lesson whose two coders had a greater than one point discrepancy were coded by a third coder, and the ratings were kept from the coders with highest agreement, providing a teacher with an overall MQI score composed of an average across two coders, across two lessons. Agreement was fair with kappa=0.38. The overall MQI scores range from 1-5, where 1 is a lesson characterized by errors, unproductive student-teacher interactions, lack of directionality, or lack of mathematics and a 5 reflects rich (in terms of explanation/justification and representations), focused lessons which include productive teacher-student interactions and mathematical practices.

Classroom Implementation of Emphasized PD Components. During year 2 and 3 of the project, we developed a 5-point rubric system to evaluate the level teachers were implementing the standards-based instruction advocated by the PD. The PD facilitators visited math classrooms of their participating teachers five times throughout the year. Based on these observations, they rated each of their teachers based on a rubric. The rubric was refined via interviews with the PD facilitators during year 2. Then, during year 3, each of the facilitators was again interviewed (after providing ratings) in order to provide evidence that their interpretations of their teachers' implementations were based on the rubric and that the rubric was being used in consistent manners across the facilitators. Each FullPD teacher has a score on a 4-point scale: a score of 0 (low) reflects low cognitive tasks with no or minimal attempts to use teaching routines and rare instances of students engaging in habits of mind and interaction. A score of 1 (low-medium) reflects low cognitive demand tasks, but use of some teaching routines (with emphasized teaching moves), and some evidence of student engagement in habits. A score of 2 (medium-high) reflects medium to high cognitive tasks, use of many teaching routines (with emphasized teaching moves), and many students engaged in habits. A score of 3 (high) is similar to 3, but requires high cognitive demand and students moving towards justifying and generalizing.

An overview of the teacher data can be found in Table 3.

Table 3

Number of Teachers and Types of Teacher Data

Data Type	FullPD	SummerPD
MKT Baseline and Year 1	63	37

MKT Baseline, Year 1, Year 2	47	19
MQI Baseline and Year 1	52	31
MQI Baseline, Year 1, Year 2	41	13
Implementation Scores	83	NA

Student Standardized Assessments. At the student level, we collected year-end student assessment scores for all third through fifth grade students for the three years of the project, and a year of baseline data prior to the project. For baseline and year 1 of the project the students took the *Measurements of Student Progress* (MSP). In year 2 and year 3, the school district switched to the *Smarter Balanced Assessment* (SBA). We use percentiles (from state level data) rather than scores to keep the assessments on the same scale and score relative to the state population. Both measures were designed to assess the extent to which students met grade level standards and included multiple-choice and short answer questions assessing procedural fluency and conceptual understanding (Office of Superintendent of Public Instruction, n.d.; 2010). While both assessments included the opportunity for students to "show how you know" (MSP) and "show how you found your answer" (SBA), the SBA provided a greater variety in the ways students could represent solutions and explanations through technology enhanced items (e.g., sort shapes, draw lines of symmetry, use a number line to show the relative size of fractions) and performance task items (e.g., accomplish tasks required of a grocery store manager opening a new store). Additionally, both measures had substantial review to ensure that all content is free of biased references to culture, ethnicity, or gender (Office of Superintendent of Public Instruction, 2012; Smarter Balanced Assessment Consortium, 2021). This data set allowed for a holistic look at the impact on students as measured by the assessments valued by the district.

3.2.3 Methods of Analysis. To answer our research questions, we conducted several types of analyses. For analysis of MKT and MQI growth, we conducted paired *t*-tests from baseline to year 1 and baseline to year 2. Additionally, we conducted a 2-factor repeated measure ANOVA on the MKT scores for teachers with a complete data set. (A more nuanced structural equation can be found in the supplementary material.)

To address the complicated student assessment data, we created cross-classified 3-level hierarchical linear models (HLM) (e.g. Hong & Raudenbush, 2008) using the R package lme4. In this model, student test scores over the years are nested within students (level 1). Students are nested within teachers (level 2), but can change teachers throughout the project. Teachers are then nested within schools (level 3). This approach fits the nature of our data where we have a range of repeated measures (students with 1, 2, or 3 tests) that can be associated with the number of treatment years and nested within students. We created several models with a range of demographic variables that have been associated with student math assessment scores (Carnoy & García 2017), focusing on eligibility for free and reduced lunch, transitional bilingual status, race, and gender. We eliminated demographic variables including gender at the student level which were not significant in any of the models (see supplementary materials).

In addition to the full model, we also created a similar model with just FullPD schools that included implementation at the teacher level.

The model equation for the test score level (level 1) is as follows:

$$Percentile_{ijkl} = \pi_{0jkl} + \pi_{1jkl} *(Years\ in\ FullPD\ Class) + \pi_{1jkl} *(Years\ in\ SummerPD\ Class) + e_{ijkl}$$

where Percentile$_{ijkl}$ represents the estimated *ith* test score (percentile) for student *j* in teacher *k*'s class in school *l*; π_{0jkl} is the random intercept for student *j* in teacher *k* and school *l*; π_{1jkl} is the slope for number of years in an FullPD classroom; π_{2jkl} is the slope for number of years in an

SummerPD classroom π_{1jkl} as of the *i*th test for student *j* in teacher *k* and school *l*; and e_{ijkl} is the level-1 random effect (residual).

The cross-classified level 2 model (student crossed with teacher) is as follows:

$$\pi_{0jkl} = \theta_{0l} + (\gamma_{01l})*BILING_j + (\gamma_{02l})*BLACK_j + (\gamma_{03l})*LATINO_j + (\gamma_{04l})*OTHERMIN_j$$
$$+(\gamma_{05l})*FREELUNCH_j \qquad +(\gamma_{06l})*BILING_j *LATINO_j + b_{00j} + c_{00k}$$

where θ_{0l} reflects the initial estimate for a test-score in school *l*; γ_{01l} is the main effect for the binary coded transitional bilingual status of student *j* ($BILING_j$); $\gamma_{02l}, \gamma_{031}, \gamma_{04l}$ are the main effects of a student being Black, Latino/Hispanic, Other minoritized (Multiracial, Native American, Pacific Islander) respectively; γ_{051} is the main effect of the binary code eligibility for free and reduced lunch of student *j* ($FREELUNCH_j$); γ_{061} is the interaction effect of being transitional bilingual and Latino/Hispanic; b_{0j} is the random effect of student *j*; and c_{00kl} random effect of teacher *k* in school *l.*

Each of the significant main effects were also included in the sum for the coefficient π_{1jkl} and π_{1jkl} to capture the interaction effect between student *j*'s background variable and test occurrence *i* number of years in a FullPD or SummerPD classroom, respectively.

$$\pi_{1jkl} = \theta_{0l} + (\gamma_{11l})*BILING_j + (\gamma_{12l})*BLACK_j + (\gamma_{13l})*LATINO_j \ (\gamma_{14l})*OTHERMIN_j$$
$$+(\gamma_{15l})*FREELUNCH_j$$
$$\pi_{2jkl} = \theta_{0l} + (\gamma_{21l})*BILING_j + (\gamma_{22l})*BLACK_j + (\gamma_{23l})*LATINO_j \ (\gamma_{24l})*OTHERMIN_j$$
$$+(\gamma_{25l})*FREELUNCH_j$$

The school level (level 3) model is as follows:

$$\theta_{0l} = \delta_{000} + d_{00l}$$

where δ_{000} is the intercept and d_{00l} is the level 3 residual.

Our second set of models (implementation models) were limited to just the FullPD schools changing our series of equations to have only *Years in FullPD* (and not *Years in SummerPD*) at the score level and the addition of $IMPLEMENTATION_k$, the main effect for the teacher *k*’s implementation score (centered) at level 2.

In order to assure that these models were appropriate, we checked several assumptions. In the full (FullPD and SummerPD model), a D'Agostino skewness test indicated a small, but significant right skewness of the Level 1 error distribution, 0.12. The Anscobe-Glynn kurtosis test reflected slightly heavy-tails (kurtosis of 4.0). For the implementation model, the error distribution was not significantly skewed right (skew=.05) with long tails (kurtosis of 4.4). Yuan and Bentler (2005) show that HLM models can be asymptotically robust to departures from normality. To verify, we created dummy variables for outliers (observations with standard residuals greater than 2). This inclusion of the dummy variables did not affect the size or direction of any other predictors in the models. To detect possible issues of multicollinearity, we used the approach advocated by Clark (2013) finding no MVIF values greater than 2.5. In table 4, 5, and 6, we provide an overview of the data in these models.

Table 4

Number of Test Occurrences disaggregated by demographics and treatment

		SummerPD Schools	FullPD Schools
Black	Free Lunch	2305	1786
	Not Free Lunch	428	349

Latino/Hispanic	Free Lunch	2247	1422
	Not Free Lunch	335	268
White or Asian	Free Lunch	3917	2562
	Not Free Lunch	3309	2579
Other	Free Lunch	755	690
	Not Free Lunch	194	169

Table 5

Number of Test Occurrences disaggregated by Latino/Hispanic, transitional bilingual status and treatment

	SummerPD Schools	FullPD Schools
Latino/Hispanic and transitional bilingual	959	597
Not Latino/Hispanic and transitional bilingual	814	596

Table 6

Number of Participating Students, Teachers, Schools

	SummerPD Schools	FullPD Schools
Level 2 - Students	7649	5850
Level 2 - Teachers	271	200
Level 3 - Schools	20	13

4 Results

We first share results related to the hypotheses around teacher changes in our model which were consistent with prior empirical work and our hypotheses. We then dedicate more substantial discussion to the student results.

4.1 Participating Teachers Across Both Treatments Saw Gains in their Mathematical Knowledge for Teaching. We measured teachers MKT at three points in time (baseline, after year 1, and after year 2.) A 2-factor ANOVA with repeated measures reflected a significant within subject effect, that is teachers made gains from year to year, ($F(2,128)=5.9$, $p<0.01$), but insignificant difference between SummerPD and FullPD teachers ($F(1,64)=0.29$, $p=0.59$) or difference in interaction between the SummerPD/FullPD and within subject effect ($F(2,128)=0.1$, $p=0.90$). The average LMT score (using the standardized scoring scale) for baseline was -0.01 ($SD=0.84$), year 1, 0.16 ($SD=0.90$), and year 2, 0.27 (0.87). The effect size, cohen’s d, from baseline to year 2 was 0.32 reflecting somewhere between a small and medium effect.

4.2 Mathematical Quality of Instruction Increased for FullPD Teachers Only. Both groups started with a overall MQI (1-5) mean score of 3.26 (SummerPD *SD*=0.54, FullPD *SD*=0.58) but after one year, SummerPD teachers stayed relatively steady (mean of 3.24, *SD*=0.68) and FullPD participants made gains to a mean of 3.47 (*SD*=0.55). The FullPD gain reflected a small to medium effect size (*d*=0.39). Due to attrition, we did not further analyze the SummerPD teachers (as the set of remaining teachers scored significantly lower on the baseline than the teachers who did not continue). However, we note the FullPD teachers did not continue to make gains, but did maintain the higher mean in year 2 (*M*=3.41; *SD*=0.62). (See supplementary material for more information.)

If we consider the results surrounding teacher change, we can point to the success of our model. We hypothesize that the extended work around standards-based instruction during the academic year (within the schools) accounts for the gains in MQI, as this instrument has a strong overlap with standards-based instruction.

4.3 The Relationship Between PD and Student Achievement

In this section, we present the two types of models: the full model (table 7 and 8) including the full set of student assessment data from the participating schools and the implementation model (table 9 and 10) including just the subset of FullPD schools whose teachers for whom we had implementation scores.

4.3.1 The Model Including FullPD and SummerPD Schools

Table 7

3-level cross-classified model with outcome percentile on year-end mathematics assessment including main effects and interactions

Fixed Effect[1]	Coefficient	Standard Error	App. d. f.	t-value	p-value
(Intercept)	56.12	1.08	44	52.03	0.00
Main Effects					
Free Lunch	-11.69	0.57	19289.90	-20.36	0.000
Black	-11.93	0.64	19468.89	-18.68	0.000
Latino	-5.24	0.75	19000.01	-6.96	0.000
OtherMinoritized	-10.11	0.92	21154.84	-11.01	0.000
Transitional Bilingual	-13.35	0.99	17909.40	-13.55	0.000
YearsFullPD	1.32	0.38	17658.06	3.49	0.000
YearsSummerPD	0.16	0.34	17106.57	0.48	0.631
Interactions					
Transitional Bilingual*Latino	-3.00	1.32	13204.90	-2.27	0.023
Black:YearsFullPD	-1.96	0.44	15075.61	-4.42	0.000
Black:YearsSummerPD	-0.73	0.40	14537.76	-1.81	0.070
Latino:YearsFullPD	-1.34	0.51	15750.93	-2.65	0.008
Latino:YearsSummerPD	-0.72	0.43	14357.00	-1.69	0.092
OtherMin:YearsFullPD	-0.76	0.66	15533.88	-1.17	0.244
OtherMin:YearsSummerPD	0.38	0.61	15023.07	0.63	0.532
BiLing:YearsFullPD	1.86	0.55	15484.23	3.40	0.001
BiLing:YearsSummerPD	1.63	0.47	13986.20	3.46	0.001
FreeLunch:YearsFullPD	-0.74	0.40	16079.48	-1.84	0.065
FreeLunch:YearsSummerPD	0.05	0.35	15200.53	0.15	0.879

1. In general, we did not include insignificant effects. Exceptions: YearsSummerPD terms for comparison to FullSummerPD; FreeLunch interaction as it was near significant and significant in some models; OtherMinoritized interactions as they trend the same direction as our general results -- although this group of students is much smaller.

Table 8

Variances of Random Effects

Source	Sample Size	Null Model	Full Model
Student	13499	543	471
Teacher	471	39	38
School	33	77	26
Level 1	23315	114	112
Total		*773*	*647*

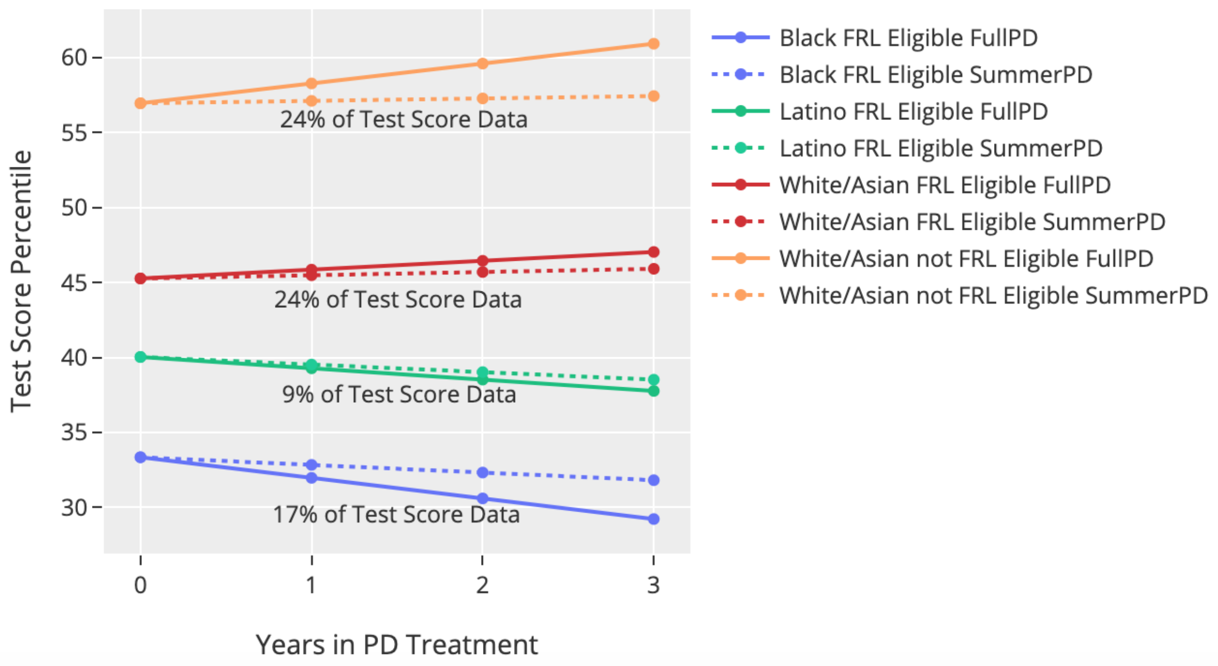
*The model is able to reduce the variation at the school level by ~ 66%, student level by ~13%.

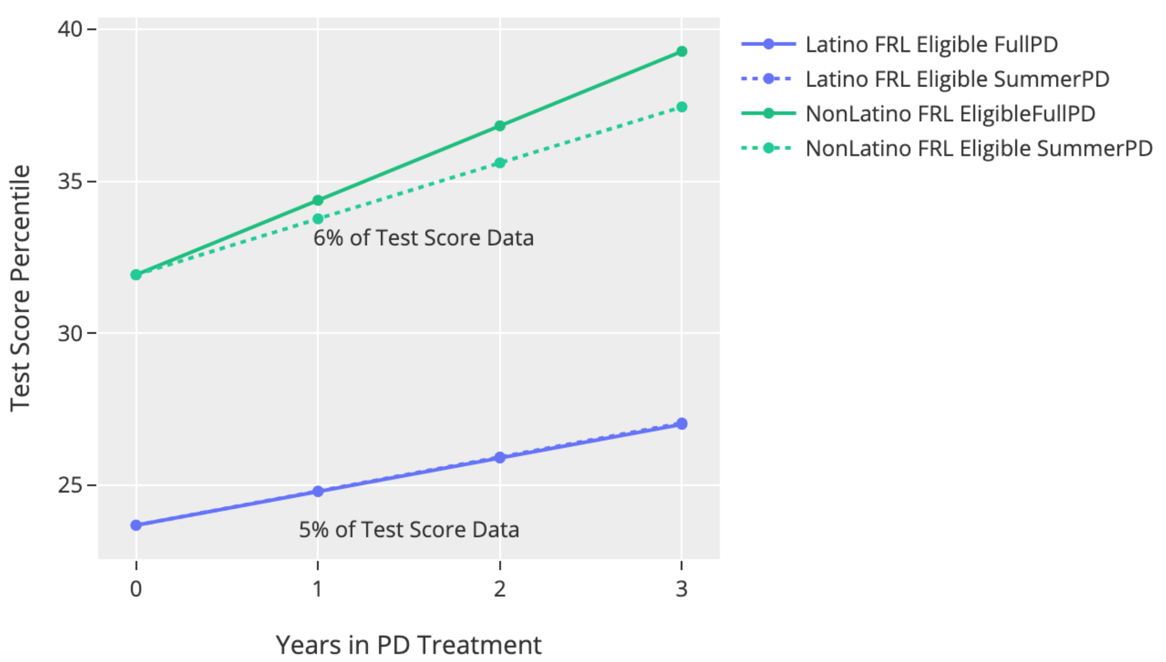
If we focus on main effects, we can see that students are making significant gains for multiple years in FullPD while students in the SummerPD are not. The effect size is rather small at 1.32 percentiles difference per year in a FullPD school -- that's an effect of 0.05[4] standard deviations or 0.14 standard deviations for three years. Although the effect is small, 43.6% of the data would have to be due to bias to invalide the inference at alpha = 0.05 (in accordance with Frank et al., 2013). Thus, we can suggest there is some effect on student performance on assessment when in the FullPD schools.

Figure 5

[4] Effect size is calculated from effect divided by standard deviation of the baseline data (28.21).

Estimated Percentile Scores for Non-Transitional Bilingual Students Over the Course of the Project Inclusive of All Demographic Categories Containing At Least 5% of Test Scores: Non-Transitional Bilinguals (top); Transitional Bilinguals (bottom)





While this is a promising result, the significant interaction terms change our interpretation. First we note that in both types of schools, transitional bilingual students saw additional estimated gains beyond just their years in the classrooms (see Table 7 / Figure 4). However, students who were Black or Latino/Hispanic, but not transitional bilingual, did not see gains. That is, the opportunity gap was widened. We also note socioeconomic status (FRL) interaction with treatment years was borderline significant. We left the effect in this model, but note because the effect is not robust (significant in some models and not others), we cannot make substantial claims related to socioeconomic status. Further, about 85% of students minoritized groups (Black, Latino/Hispanic, Other) were eligible for FRL limiting the ability to unpack the intersectionality of race/ethnicity and socioeconomic status.

In contrast, race/ethnicity, and in particular whether a student was Black, had a significant negative interaction with years in FullPD in all model variations. This finding was robust. In order to invalidate the inference (at alpha=0.05), 56% of the estimate would have to be due to bias. That is, to invalidate an inference, 13,056 exam scores would have to be replaced with cases for which the effect is 0. To contextualize, consider the growth lines in Figure 5. A student who is Black and eligible for FRL is estimated to decrease 1.38[5] percentiles for each year in a FullPD school or 4.14 percentiles over three years -- a small effect of 0.15 standard deviations. In the same school, a White or Asian student who is not eligible for FRL would be estimated to increase by 1.32 percentiles per year, or 3.96 percentiles over three years.

If we compare FullPD to the SummerPD schools, we can observe this widening opportunity gap does not exist to the same degree in SummerPD schools. This is also reflected in

[5] Calculated: Black*YearsinFullPD+YearsinFullPD+FRL*Black*YearsinFullPD

the lack of significance of the interaction effects in the model. We can also see Latino/Hispanic (FRL eligible) and Asian/White students (FRL eligible) see parallel growth, respectively, in SummerPD and FullPD. In both contexts, the Latino/Hispanic students were estimated to make very small declines, and Asian/White students were estimated to make very small gains. In this way, the FullPD did not appear to amplify opportunity gaps for these students, but also did not disrupt them.

4.3.2 Looking at the Effect of Teacher Implementation in FullPD Schools. In order to further understand these results and relationships to our study assumptions, we consider a second model that only includes FullPD schools to examine how teacher implementation of the standards-based instruction relates to student outcomes. (Table 9 & 10)

Table 9

3-level cross-classified model for FullPD schools only with outcome percentile on year-end mathematics assessment and addition of implementation variable

cluding main effects and interactions

Fixed Effect	Coefficient	Standard Error	App. d. f.	t-value	p-value
(Intercept)	59.37	2.03	17.21	29.26	0.000
Main Effects					
Free Lunch	-12.49	1.18	5242.68	-10.63	0.000
Black	-11.92	1.38	5230.68	-8.66	0.000
Latino	-5.80	1.64	5280.93	-3.54	0.000
OtherMinoritized	-10.72	1.99	4891.23	-5.40	0.000

Transitional Bilingual	-10.06	1.74	3933.15	-5.77	0.000
YearsFullPD	0.93	0.52	3276.59	1.78	0.075
Implementation	1.81	0.74	61.20	2.45	0.017
Interactions					
Transitional Bilingual*Latino	-10.06	1.74	3933.15	-5.77	0.000
Black:YearsFullPD	-2.50	0.74	2378.01	-3.37	0.001
Latino:YearsFullPD	-1.31	0.87	2457.18	-1.51	0.131
Other:YearsFullPD	-0.75	1.18	2304.03	-0.64	0.525
BiLing:YearsFullPD	-2.40	2.62	3938.90	-0.91	0.361
FreeLunch:YearsFullPD	0.11	0.62	2441.05	0.17	0.862

Table 10

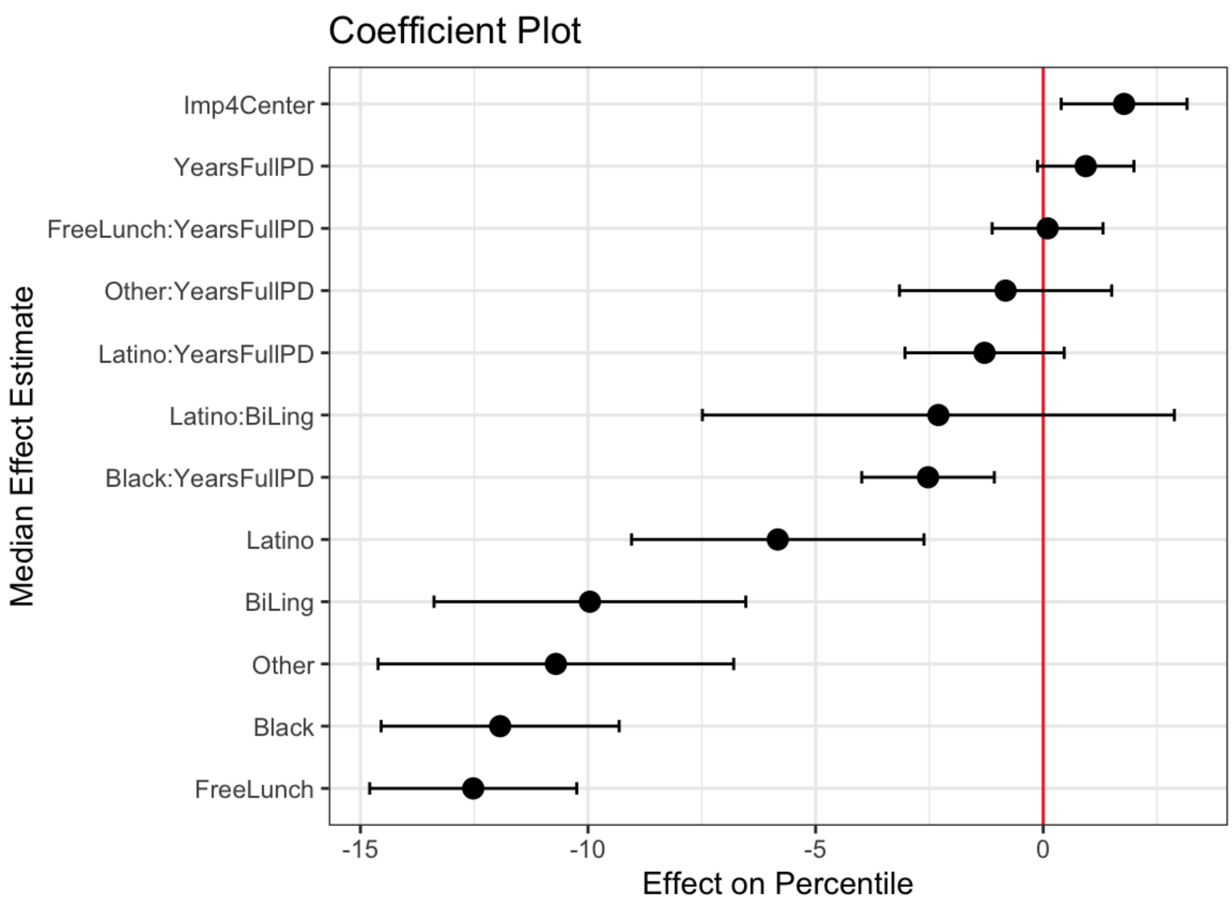
Variances of Random Effects

Source	Sample Size	Null Model	Model 2
Student	4004	557	488
Teacher	83	44	34
School	13	113	33
Level 1	5366	107	107
Total		*821*	*662*

This model includes an implementation variable at the teacher level (0-3 score as reported by the PD facilitators). The implementation variable is centered (*Mean*=1.37) and provides more insight into the degree to which teachers were incorporating the standards-based approach into their classroom teaching. We note that there are less test scores (*n*=5,399) due to the inclusion of a teacher level variable (leaving a model for only consenting teachers at FullPD schools). This reduces our power, but the directionality of nearly all the effects remain consistent with the exception of the interaction between socioeconomic status (FRL) and years of PD. Additional analysis did not point to substantial differences between the full data set and the data set from consenting teachers, and thus, we suggest this model provides insight into the larger project. As in the prior model, and in alignment with general trends, race and ethnicity, language, and socioeconomic status still account for an outsized effect on assessment scores (see Figure 6).

Figure 6

Coefficient plot reflecting the size of the effect of different variables on student assessment scores



Implementing the standards-based instruction is significantly and positively related to estimated scores. The estimated test score for a student classroom where a teacher incorporates most of the PD’s emphasized approach to instruction (score=3) is estimated to perform 5.43 percentiles higher than if they were in a classroom of a teacher implementing limited standards-based instruction (score=0). This is equivalent to 0.19 standard deviations: a small, but notable effect. These results provide additional evidence to the larger literature base that a standards-based approach to instruction can link to student gains in assessments.

However, we found no significant interaction between race and implementation (Black, $p=0.24$; Latino, $p=0.85$, see supplementary materials). Further, the model including the implementation variable did not substantially change the interaction terms between Black and Latino/Hispanic and years in the FullPD schools. While the effect for Latino/Hispanics is no longer significant ($p=0.06$), the general trend from the more complete model stayed consistent. We can interpret this result as the degree of implementing standards-based instruction neither amplified nor attenuated the increasing opportunity gap. Implementing standards-based instruction is estimated to increase all students' scores, but is not explaining this difference in estimated gains by race/ethnicity.

4.3.3 Discussion of Student Achievement Analysis. At a basic level, we can claim that student assessment performance increased for some students in multiple years in the FullPD classrooms. We also documented that teacher implementation of the instructional model had a significant, positive correlation with student achievement scores. This result further evidences the potential for standards-based instruction to positively impact students. However, Latino/Hispanic and Black students did not see the same gains as their peers in FullPD classrooms -- and the opportunity gap for Black students was wider in the FullPD than SummerPD.

5 Discussion

In many ways, our study played out as intended: all teachers made gains in their mathematical knowledge for teaching, but only teachers in the extended model (aligned with best practices in PD) demonstrated gains in their mathematical quality of instruction. We also documented that teachers implementing our instructional model with higher fidelity had students with higher assessment scores, and the main effect of years in the FullPD was positive. However, our last hypothesis failed-- the PD did not lead to more equitable outcomes. While we found some positive results (e.g., emerging bilingual students attained positive gains above and beyond their peers), Black (and to some degree Latino/Hispanic) students were left behind, pointing to a biased experience we had not anticipated.

5.1 Making Sense of Inequitable Estimated Achievement Scores

In order to make sense of this result, we consider possible explanations for how our underlying theory of action did not play out as intended.

5.1.1 Explanation: Fidelity of Implementation and Time for Change in Practice. When considering the impact of a PD program, the fidelity of its implementation is an essential component. If the PD is not being implemented as intended, then the hypotheses are irrelevant. However, in our setting, we have reason to believe the PD program was taken on with a relatively high degree of fidelity. All schools worked with experienced PD facilitators who met regularly throughout the duration of the project. School leaders were also incorporated into all aspects of the PD, with principals and coaches trained to observe classrooms using a protocol that emphasizes the PD’s instructional vision. Additionally, project team members observed several project classrooms, taking detailed field notes related to how the PD was implemented. As such, it is unlikely that our results can be explained by a program fidelity issue, although we acknowledge that more systematic evidence of implementation could strengthen this claim.

Another explanation could be the lack of implementation by teachers at the classroom level. During a three year long PD, we observed many teachers only beginning to make shifts in their classroom practice, with some teachers making no change at all (Fagan et al., 2017). From this perspective, the student assessment data may reflect that teachers were still in the process of change, and that some of the equitable teaching practices had not yet fully materialized. Our implementation scores did correlate with higher assessment scores, but did not mediate inequitable effects. It may be the case that these scores are not sensitive enough to implementing standards-based practices in equitable ways.

5.1.2 Explanation: The PD did not have the content and focus needed to equitably impact student outcomes. As Barajas-López and Larnell (2019) cautioned, “An under discussed danger of equity discourse within mathematics education is that such discourse can also be too easily re-rendered as race-neutral—or, much more often, as race-lite” (p. 354). That is, we often attend to inequity in ways that do not center their societal sources. There is a difference between “all students are capable of doing rich mathematics" and “Black students are capable of doing rich mathematics.” The second version explicates the counter narrative to the conditions in the United States that have limited opportunities for African Americans and often positioned Black students as academically inferior (Ladson-Billings, 1997). Battey and Leyva (2016) further highlighted “whiteness” or “the institutional ways in which white supremacy in mathematics education acts to reproduce subordination and advantage” (p. 51). We use their framing, and the work of other researchers, to explore ways that the FullPD may have served to unintentionally reproduce societal inequities.

This PD focused on many important aspects of promoting more equitable classrooms: attending to issues of status amongst students, positioning students as the mathematical authority, and engaging students in rich tasks and discussions focused on conceptual understanding. However, this approach was not focussed on anti-racist or anti-bias education. Despite many of the schools in our study being majority-minority, the fact that students were Black and/or Latino/Hispanic was not a focus of discussion. To evidence this, we did a text search for “Black” or “African American” across three years of field notes related to the PD in one of the majority-minority schools. These notes covered a total of 30 days of PD discussions and were created to be as close to live transcribing the conversations as possible. Neither term came up in any of the discussions amongst teachers and PD facilitators. This is not surprising as both the teachers in the schools, and the PD facilitator and designers, have likely been trained in a “color blind” fashion, meaning race is not discussed or acknowledged (Glazier, 2003), and conversations around race are often uncomfortable (Gordon, 2005). Furthermore, the majority of the PD facilitators and 80% of the teachers in the FullPD schools were white. Without space to address this potential misalignment between teacher/facilitators and students’ experiences, it is unlikely participants are positioned to become more aware of how whiteness may be unintentionally reinforced.

Battey and Leyva (2016) identified a number of dominant yet problematic ideological discourses including treating mathematics as a “cultureless” domain, propagating racial hierarchies (Martin, 2009) in terms of ability, and treating ability as innate. The PD facilitators explicitly worked with teachers to move away from mathematical ability as an “innate” characteristic. However, the tools and emphases of the PD never challenged mathematics as a cultureless domain or directly addressed any racial bias that may exist in relation to academic ability. By leaving these aspects implicit, we did not disrupt prevalent discourses such as racial stereotypes (Picower, 2009) that support and maintain whiteness. Further, although the PD emphasized positioning students as the mathematical authority, students who participate in ways that are viewed as “white” may hold more authority and have greater access to content over their peers (Battey & Leyva, 2016).

Battey and Leyva (2016) also addressed the important role of maintaining high cognitive demand for all students, a shared theme throughout the PD. However, incorporating high cognitive demand tasks does not guarantee all students have equal access to these opportunities. Who contributes what type of mathematics in the classroom can easily be shaped by implicit bias (Reinholz & Shah, 2018) and status issues that emerge as students interact with each other and the teacher (Bartell, 2011). Although peer-to-peer status was explicitly addressed in the PD, the idea that status in the classroom may reproduce larger racial narratives about competency was never addressed. To disrupt status hierarchy informed by race, assigning competence should extend beyond use as a general tool and be used intentionally to amplify the mathematical contributions and abilities of minoritized students.

Several qualitative studies have pointed to standards-based instruction as insufficient for more equitable outcomes, particularly when considering the experiences of Black students. For example, Murrell’s (1994) study of Black male students in middle school mathematics classrooms points to a variety of ways in which these students participated and understood the classroom that differed from their peers, such as interpreting teacher feedback as a commentary on the quality of their performance rather than directed at conceptual learning. Wilson et al.

(2019) pointed to specific ways that conceptually-oriented classrooms may or may not support Black students on test achievement. Black students demonstrated greater achievement in classrooms that were characterized by making mathematical expectations explicit, coaching students socially and mathematically, attending to local context (both in service of and not in service of mathematics), attending to language, delegating mathematical authority to students, positioning individuals as competent, and attending to the classroom community. When we contrast this work with the focus of the PD, there is some positive overlap (delegating mathematical authority to students), yet some areas could have been attended to, including positioning individuals (rather than students collectively) as competent, and explicitly focusing on students’ language/background as assets. These studies point to the insufficient nature of a “one-size fits all” approach to standards-based instruction.

5.1.3 Explanation: Underestimating the complexity involved in the classroom and a need to move beyond just social constructivism. Our underlying change model focused on components of the PD, the actors involved, and relationships between teacher constructs and students. However, we overlooked the surrounding cultures and environment. Louie’s (2017) case studies of equity-minded teachers in an urban high school illustrated that even with complex instruction training, teachers often reproduced a culture of exclusion driven by a view of low and higher performing students that permeates mathematics education. Further, a single teacher’s impact is dwarfed by many factors beyond their control, from school resources to societal narratives about race to socio-economic inequities. In their analysis of socioemotional outcomes and intersectionality, Bécares and Priest (2015) argued that, “Interventions to eliminate achievement gaps cannot fully succeed as long as social stratification caused by gender and racial discrimination is not addressed” (p. 13). Larger societal narratives alter the way that classrooms are experienced and are not easily disrupted (Battey & Leyva, 2016). For example, managing stereotype threat is emotionally taxing on students, and because stereotype threat exists outside the classroom, it will continue to impact experiences within the classroom.

The FullPD was tied to teachers’ practice and focused on their students’ mathematical thinking. This is an important step in centering students as contributors to mathematics, but this alone does not account for how students’ respective cultures and backgrounds shape their experiences within a student-centered classroom. Further, disrupting narratives around who is capable of doing high-level mathematics may require support beyond discussing mindset and complex instruction. We entered this project with a theory of mathematics learning driven by social constructivism. However, we now recognize such theories need to be paired with other theories that acknowledge the role of race, bias, and status in society -- “theoretical perspectives that see knowledge, power, and identity as interwoven and arising from (and constituted within) social discourses” (Gutiérrez, 2013, p. 40). Such a stance has been well-articulated and argued for through empirical examination (e.g., Berry et al. (2014)’s critical race theory informed analysis of the literature). We hope our results contribute to this discourse and provide further support to move beyond a “mathematics for all” image to integrating culturally responsive pedagogy and explicit attention to anti-bias and anti-racism instruction.

5.2 Limitations and Directions for Future Research

First, our results only reflect one study in one particular school district. While we evidence some important findings, we suggest replicability in other settings. Second, we must acknowledge that all the results presented in this study hinge on the instruments used. We leveraged well-studied instruments (LMT; MQI) that reflect meaningful components of the PD. We conjectured that as teachers engaged more with student thinking, their MKT would increase.

However, we limited our knowledge exploration to number and operation which is not the only domain relevant to elementary mathematics-- and not the only mathematical domain relevant to the PD sessions. In our findings, we also documented FullPD teachers' change in practice as measured by the MQI. The MQI positions mathematics neutrally and does not address important aspects of equity such as who participates (Charalambous & Litke, 2018). For example, a teacher could have an overall score of 5 (the highest level) for a lesson in which only white students participated. Future research could use a tool such as EQUIP (Reinholz & Shah, 2018) to provide more insight into who is contributing what type of mathematics, and how participation rates for particular groups of students (based on race, gender, or other classification) compare to the class composition. Other tools, such as Aguirre and del Rosario Zavala’s (2013) lesson analysis tool or Waddell’s (2014) culturally ambitious teaching practices observation framework could further illustrate important classroom aspects that focus on mathematical thinking, language, culture, and/or social justice.

For student outcomes, we relied on standardized assessment data. As argued by Gutiérrez (2012), performance on assessments is a measure of how well students can “play the game” that reflects a dominant culture. While we acknowledge the limits of looking at equity from a purely achievement-based stance, achievement equity is important in society, and gives us a lens to recognize limitations in our work and assumptions. Additionally, the coarseness of student assessment data does not fully capture all the elements of learning from the PD. Further, the assessment changed during our study. However, both assessments measured conceptual and procedural knowledge and were evaluated for race/ethnicity/gender bias. We did not see evidence that SBA was less aligned with the PD’s focus, and so is unlikely to explain our inequitable results. We also acknowledge that our model only included 3rd-5th grader assessments, meaning no single student had baseline, year 1, year 2, and year 3 data. The cross-classified approach allowed us to make the best use of the test scores available, but these estimates will include errors based on the nature of student data in the school setting. However, a substantial amount of bias would be needed to invalidate our results.

Moving forward, we aim to empirically explore data from our classroom videos to better make sense of this achievement inequity by identifying classrooms in which Black and Latino/Hispanic students outperformed predictions and classrooms with achievement outcomes that aligned with the model. From these classrooms we can address whether there is a difference in how these teachers implemented components of the PD as outlined in the FullPD Framework (Melhuish & Thanheiser, 2017). This would lend credence and nuance to the first explanation of our results. Further, we plan to explore participatory equity and how Black and Latino/Hispanic students are positioned by teachers and each other. Through analyzing these classroom components, we hope to develop more robust hypotheses about the achievement results.

We would suggest future, large-scale, quantitative studies to incorporate data collection to allow for systematically using tools such as EQUIP to complement achievement equity analysis with participation equity analysis. Additionally, the role of socioeconomic status was not clear cut in our analysis, and additional studies are needed to clarify its role.

5.3 Transforming an Effective for Some PD to be Effective for All

Our results suggest that the PD model in the study was effective in engendering much of the change we anticipated. The model provides space for teachers to collaboratively plan, practice, and debrief. The PD also introduced a number of tools to focus this planning, teaching, and reflection. The teachers and school leaders, along with the PD facilitator, begin with an existing curriculum lesson and identify the core conceptual understanding, alter existing tasks to increase cognitive demand, and anticipate student thinking and ways to support students in high level engagement through teaching routines. After a lesson is enacted, they reflect on the types of contributions students made, the types of teacher moves that seemed supportive, and at some points, reflect on status in their classrooms and whether there are ways to lift particular voices.

However, the standards-based approach did not explicitly attend to racial bias and the experiences of the particular students in a classroom was insufficient. Rubel (2017)'s recent look at pedagogy highlights the importance of multiple types of equity-based instructional practice and the inherent difficulty of incorporating practices from culturally relevant pedagogy and teaching mathematics for social justice. Further, Martin (2007) argues for teaching components necessary for teachers to work with racially diverse students including "developing deep understanding of the social realities experienced by these students" (p. 25).

Future iterations of the PD[6] will incorporate bias reduction as a major component. A first step towards that is the inclusion of the *Social Justice Standards: The Teaching Tolerance Anti-Bias Framework* (SJS) (Chiariello et al., 2016). The SJS provide a common language to guide teaching, increase understanding of difference, and to actively challenge bias, stereotyping and all forms of discrimination in schools and communities. The PD will also include reflection on participatory equity and whose ideas are raised (moving beyond a surface examination of whose ideas come to the surface) as well as connecting to students and community by creating high demand tasks that connect to students' lives, and the community. This new focus will be supported through existing tools that promote equitable instruction beyond just a standards-based

[6] NSF DRL- 2101665 "Developing and Researching K-12 TLs Enacting Anti-Bias Mathematics Education (DREAM)" http://cadrek12.org/projects/developing-and-researching-k-12-teacher-leaders-enacting-anti-bias-mathematics-education-0

approach (e.g., Aguirre & del Rosario Zavala, 2013; Moschkovich, 2013) and center equitable teaching practices within our existing frameworks.

5.4 Conclusion

In this paper, we presented a narrative of well-designed professional development built directly from literature about effective PD. The intervention contained a set of important components linked to engendering teacher change and ultimately promoting a student-centered, standards-based classroom. In many ways, this PD was a success; we saw gains in teacher knowledge, gains in quality of instruction (for FullPD teachers), and some impact on student standardized assessments. However, this work also highlighted that minoritized students did not see the same gains as their counterparts. The PD was informed directly by work from Railside (Boaler & Staples, 2008) and the QUASAR project (Silver & Stein, 1996) that have documented equitable outcomes from a standards-based approach. Yet, the classrooms in our project did not reflect similar results.

We can make many conjectures for why this is the case, but the most obvious is that we entered the project with flawed assumptions. The NCTM Research Committee (2017) has challenged mathematics education researchers to re-evaluate our work through the lens of power, treating our work as inherently political to “help us interrogate our views and assumptions and understand our experiences and choices differently” (p. 127). We have re-evaluated our assumptions that a standards-based approach to PD can be successfully implemented without attention to race and the particular backgrounds of the students involved. We position our study as cautionary to educators who may take a similar approach to research-informed and standards-based PD; “mathematics for all” may not be mathematics for all.

References

Aguirre, J. M., & del Rosario Zavala, M. (2013). Making culturally responsive mathematics teaching explicit: A lesson analysis tool. *Pedagogies: An international journal*, *8*(2), 163-190.

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond, & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice*, (pp. 3-22). Jossey-Bass Publishers.

Barajas-López, F., & Larnell, G. V. (2019). Research Commentary: Unpacking the Links Between Equitable Teaching Practices and Standards for Mathematical Practice: Equity for Whom and Under What Conditions?. *Journal for Research in Mathematics Education*, *50*(4), 349-361.

Bartell, T. G. (2011). Caring, race, culture, and power: A research synthesis toward supporting mathematics teachers in caring with awareness. *Journal of Urban Mathematics Education*, *4*(1), 50-74.

Battey, D., & Leyva, L. A. (2016). A Framework for Understanding Whiteness in Mathematics Education. *Journal of Urban Mathematics Education*, *9*(2), 49-80.

Bécares, L., & Priest N (2015) Understanding the Influence of Race/Ethnicity, Gender, and Class on Inequalities in Academic and Non-Academic Outcomes among Eighth-Grade Students: Findings from an Intersectionality Approach. *PLoS ONE 10*(10): e0141363. https://doi.org/10.1371/journal.pone.0141363

Bell, C. A., Wilson, S., Higgins, T., & McCoach, D. B. (2010). Measuring the effects of professional development on teacher knowledge: The case of developing mathematical ideas. *Journal for Research in Mathematics Education*, *41*(5), 479-512.

Berry III, R. Q., Ellis, M., & Hughes, S. (2014). Examining a history of failed reforms and recent stories of success: Mathematics education and Black learners of mathematics in the United States. *Race Ethnicity and Education*, *17*(4), 540-568.

Blank, R. K., De Las Alas, N., & Smith, C. (2007). *Analysis of the quality of professional development programs for mathematics and science teachers: Findings from a cross-state study.* Council of Chief State School Officers.

Boaler, J. (2006). How a detracked mathematics approach promoted respect, responsibility, and high achievement. *Theory into Practice*, *45*(1), 40-46.

Boaler, J., & Staples, M. (2008). Creating mathematical futures through an equitable teaching approach: The case of Railside School. *Teachers College Record*, *110*(3), 608-645.

Boston, M. D., & Smith, M. S. (2009). Transforming secondary mathematics teaching: Increasing the cognitive demands of instructional tasks used in teachers' classrooms. *Journal for Research in Mathematics Education, 40*(2), 119-156.

Campbell, P. F., & Malkus, N. N. (2011). The impact of elementary mathematics coaches on student achievement. *The Elementary School Journal, 111*(3), 430-454.

Carnoy, M., & García, E. (2017). Five Key Trends in U.S. Student Performance: Progress by Blacks and Hispanics, the Takeoff of Asians, the Stall of Non-English Speakers, the Persistence of Socioeconomic Gaps, and the Damaging Effect of Highly Segregated Schools. In *Economic Policy Institute*. Economic Policy Institute. https://eric.ed.gov/?id=ED588043

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C. P., & Loef, M. (1989). Using knowledge of children’s mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal*, *26*(4), 499-531.

Charalambous, C. Y., & Litke, E. (2018). Studying instructional quality by using a content-specific lens: the case of the Mathematical Quality of Instruction framework. *ZDM*, *50*(3), 445-460.

Chiariello, E., Olesen Edwards, J., Owen, N., Ronk, T., & Wicht, S. (2016). *Social Justice Standards: The Teaching Tolerance Anti-Bias Framework.* Southern Poverty Law Center. https://www.learningforjustice.org/sites/default/files/2020-09/TT-Social-Justice-Standards-Anti-bias-framework-2020.pdf

Clark, P. C. (2013). *The effects of multicollinearity in multilevel models* [Doctoral dissertation, Wright State University]. http://rave.ohiolink.edu/etdc/view?acc_num=wright1375956788

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, *18*(8), 947-967.

Cohen, D. K., & Hill, H. C. (2001). *Learning policy.* Yale University Press.

Cohen, E. G., Lotan, R. A., Scarloss, B. A., & Arellano, A. R. (1999). Complex instruction: Equity in cooperative learning classrooms. *Theory into Practice*, *38*(2), 80-86.

Corcoran, T., McVay, S., & Riordan, K. (2003). *Getting It Right: The MISE Approach to Professional Development.* (CPRE Research Report Series. RR-055). Consortium for Policy Research in Education.

Darling-Hammond, L., & Richardson, N. (2009). Research review/teacher learning: What matters. *Educational Leadership*, *66*(5), 46-53.

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

Desimone, L. M. (2009). Improving impact studies of teachers’ professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181-199.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers’ instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis, 24*(2), 81-112.

Dweck, C. S. (2007). The secret to raising smart kids. *Scientific American Mind*, *18*(6), 36-43.

Esmonde, I., Brodie, K., Dookie, L., & Takeuchi, M. (2009). *Social identities and opportunities to learn: Student perspectives on group work in an urban mathematics classroom.*

Fagan, J., Melhuish, K. M., Thanheiser, E., Fasteen, J., Guyot, L., & Rosencrans, B. (2017). Connecting teachers’ buy-into professional development with classroom habits and practices. In E. Galindo, & J. Newton (Eds.), *Proceedings of the 39th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 459-462).

Fernandez, C. (2002). Learning from Japanese approaches to professional development: The case of lesson study. *Journal of Teacher Education*, *53*(5), 393-405.

Frank, K. A., Maroulis, S., Duong, M., & Kelcey, B. (2013). What would it take to change an inference? Using Rubin's causal model to interpret the robustness of causal inferences. *Education, Evaluation and Policy Analysis, 35*(4) 437-460.

Fullan, M. (2007). Change the terms for teacher learning. *The Learning Professional*, *28*(3), 35.

Garet, M. S., Birman, B. F., Porter, A. C., Desimone, L., & Herman, R. (1999). *Designing Effective Professional Development: Lessons from the Eisenhower Program [and] Technical Appendices*. ED Pub P.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915-945.

Glazier, J. (2003). Moving closer to speaking the unspeakable: White teachers talking about race. *Teacher Education Quarterly, 30*(1), 73-94

Goldsmith, L. T., Doerr, H. M., & Lewis, C. C. (2013). Mathematics teachers’ learning: a conceptual framework and synthesis of research. *Journal of Mathematics Teacher Education, 17*(1), 5–36.

Gordon, J. (2005). Inadvertent complicity: Colorblindness in teacher education. *Educational Studies*, *38*(2), 135-153.

Gutiérrez, R. (2012). Context Matters: How Should We Conceptualize Equity in Mathematics Education? In B. Herbel-Eisenmann, J. Choppin, D. Wagner, & D. Pimm (Eds.), *Equity in Discourse for Mathematics Education: Theories, Practices, and Policies* (pp. 17–33). Springer Netherlands.

Gutiérrez, R. (2013). The sociopolitical turn in mathematics education research. *Journal for Research in Mathematics Education*, *44*(1), 37-68.

Hill, H. C. (2010). The nature and predictors of elementary teachers' mathematical knowledge for teaching. *Journal for Research in Mathematics Education*, *41*(5), 513-545.

Hill, H. C. (2014). Mathematical Quality of Instruction (MQI): 4-point version. *Ann Arbor, MI: University of Michigan Learning Mathematics for Teaching Project*.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers’ mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371-406.

Hill, H. C., Schilling, S. G., & Ball, D. (2004). Developing measures of teachers’ mathematics knowledge for teaching. *Elementary School Journal, 105* (1), 11-30.

Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48(3), 794-831.

Hong, G., & Raudenbush, S. W. (2008). Causal inference for time-varying instructional treatments. *Journal of Educational and Behavioral Statistics*, *33*(3), 333-362.

Jackson, K., & Cobb, P. (2010, April 30 - May 4). *Refining a vision of ambitious mathematics instruction to address issues of equity* [Paper presentation]. American Educational Research Association, Denver, CO, United States.

Jacob, R., Hill, H., & Corey, D. (2017). The impact of a professional development program on teachers' mathematical knowledge for teaching, instruction, and student achievement. *Journal of Research on Educational Effectiveness*, *10*(2), 379-407.

Jacobs, V. R., Lamb, L. L., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education*, *41*(2), 169-202.

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research, 86*(4), 945-980.

Knapp, M. S. (2003). Chapter 4: Professional development as a policy pathway. *Review of Research in Education, 27*(1), 109-157.

Koellner, K., & Jacobs, J. (2015). Distinguishing models of professional development: The case of an adaptive model’s impact on teachers’ knowledge, instruction, and student achievement. *Journal of Teacher Education*, *66*(1), 51-67.

Kraft, M. A., & Hill, H. C. (2018). Developing ambitious mathematics instruction through web-based coaching: A randomized field trial. *American Educational Research Journal, 57*(6), 2378-2414.

Ladson-Billings, G. (1997). It doesn't add up: African American students' mathematics achievement. *Journal for Research in Mathematics education*, *28*(6), 697-708.

Lewis, C., & Perry, R. (2017). Lesson study to scale up research-based knowledge: A randomized, controlled trial of fractions learning. *Journal for Research in Mathematics Education*, *48*(3), 261-299.

Louie, N. L. (2017). The culture of exclusion in mathematics education and its persistence in equity-oriented teaching. *Journal for Research in Mathematics Education*, *48*(5), 488-519.

Louis, K. S., Marks, H. M., & Kruse, S. (1996). Teachers’ professional community in restructuring schools. *American Educational Research Journal*, *33*(4), 757-798.

Lubienski, S. T. (2002). Research, reform, and equity in US mathematics education. *Mathematical thinking and learning*, *4*(2-3), 103-125.

Marek, E. A., & Methven, S. B. (1991). Effects of the learning cycle upon student and classroom teacher performance. *Journal of Research in Science Teaching, 28*(1), 41-53

Martin, D. B. (2007). Beyond missionaries or cannibals: Who should teach mathematics to African American children?. *The High School Journal*, *91*(1), 6-28.

Martin, D. B. (2009). Researching race in mathematics education. *Teachers College Record*, 111(2), 295-338.

Melhuish, K., & Thanheiser, E. (2017). Leveraging classroom observation to support teachers in increasing student reasoning. In L. West & M. Boston (Eds.), *Annual perspectives in mathematics education 2017: Reflective and collaborative processes to improve mathematics teaching* (pp. 183 – 200). National Council of Teachers of Mathematics.

Melhuish, K., Thanheiser, E., & Guyot, L. (2020). Elementary school teachers’ noticing of essential mathematical reasoning forms: justification and generalization. *Journal of Mathematics Teacher Education*, 23(1), 35-67.

Michaels, S., & O’Connor, C. (2015). Conceptualizing Talk Moves as Tools: Professional Development Approaches for Academically Productive Discussions. In Resnick L., Asterhan C., & Clarke S. (Eds.), *Socializing Intelligence Through Academic Talk and Dialogue* (pp. 347-362). American Educational Research Association.

Moschkovich, J. (2013). Principles and guidelines for equitable mathematics teaching practices and materials for English language learners. *Journal of Urban Mathematics Education*, *6*(1), 45-57.

Moyer-Packenham, P. S., & Westenskow, A. (2012). Processes and Pathways: How Do Mathematics and Science Partnerships Measure and Promote Growth in Teacher Content Knowledge?. *School Science and Mathematics, 112*(3), 133-146.

Murrell, P. C. (1994). In search of responsive teaching for African American males: An investigation of students' experiences of middle school mathematics curriculum. *The Journal of Negro Education*, *63*(4), 556-569.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. NCTM

National Council of Teachers of Mathematics. (2014). *Principles to Actions: Ensuring Mathematical Success for All*. NCTM

NCTM Research Committee. (2017). Equity within mathematics education research as a political act: Moving from choice to intentional collective professional responsibility. *Journal for Research in Mathematics Education*, *48*(2), 124-147.

Neufeld, B., & Roper, D. (2003). *Coaching: A strategy for developing instructional capacity.* Annenberg Institute.

Office of Superintendent of Public Instruction (n.d.). *Mathematics Assessment*. Retrieved July 21, 2021, from https://www.k12.wa.us/student-success/testing/state-testing-overview/washington-state-smarter-balanced-assessment-consortium/mathematics-assessment

Office of Superintendent of Public Instruction. (2010, September). *Mathematics Assessment Updates for 2011*. http://media.bethelsd.org/home/programs/assessment/documents/state_assessments/math_3_5.pdf

Office of Superintendent of Public Instruction. (2012, February). *Washington Comprehensive Assessment Program Technical Report.* https://www.k12.wa.us/sites/default/files/public/assessment/pubdocs/wcap2011springadministrationtechnicalreport.pdf

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, *44*(4), 921-958.

Picower, B. (2009). The unexamined whiteness of teaching: How white teachers maintain and enact dominant racial ideologies. *Race Ethnicity and Education*, *12*(2), 197-215.

Reinholz, D. L., & Shah, N. (2018). Equity analytics: A methodological approach for quantifying participation patterns in mathematics classroom discourse. *Journal for Research in Mathematics Education*, *49*(2), 140-177.

Riordan, J. E., & Noyce, P. E. (2001). The impact of two standards-based mathematics curricula on student achievement in Massachusetts. *Journal for Research in Mathematics Education, 32*(4), 368-398.

Rubel, L. H. (2017). Equity-Directed Instructional Practices: Beyond the Dominant Perspective. *Journal of Urban Mathematics Education*, *10*(2), 66-105.

Saxe, G., Gearhart, M., & Nasir, N. S. (2001). Enhancing students’ understanding of Mathematics: a study of three contrasting approaches to professional support. *Journal of Mathematics Teacher Education 4*(1), 55-79.

Schilling, S. G., Blunk, M., & Hill, H. C. (2007). Test validation and the MKT measures: Generalizations and conclusions. *Measurement*, *5*(2-3), 118-128.

Sherin, M., & van Es, E. (2005). Using video to support teachers’ ability to notice classroom interactions. *Journal of Technology and Teacher Education*, *13*(3), 475-491.

Silver, E. A., & Stein, M. K. (1996). The QUASAR project: The" revolution of the possible" in mathematics instructional reform in urban middle schools. *Urban Education*, *30*(4), 476-521.

Smarter Balanced Assessment Consortium. (2021, January 11). *Bias and Sensitivity Guidelines.* https://portal.smarterbalanced.org/library/en/bias-and-sensitivity-guidelines.pdf

Snow-Renner, R., & Lauer, P. (2005). Professional development analysis. *Mid-Continent Research for Education and Learning (McREL)*.

Staples, M. (2007). Supporting whole-class collaborative inquiry in a secondary mathematics classroom. *Cognition and Instruction*, *25*(2-3), 161-217.

Staples, M. E., & Truxaw, M. P. (2010). The Mathematics Learning Discourse project: Fostering higher order thinking and academic language in urban mathematics classrooms. *Journal of Urban Mathematics Education*, *3*(1), 27-56.

Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of Educational Change*, *7*(4), 221-258.

Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 37*(9), 963-980.

Sztajn, P., Borko, H., & Smith, T. (2017). Research on mathematics professional development. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 213–243). National Council of Teachers of Mathematics.

Toole, J. C., & Louis, K. S. (2002). The role of professional learning communities in international education. In K. A. Leithwood & P. Hallinger (Eds.), *Second international handbook of educational leadership and administration* (pp. 245-279). Springer, Dordrecht.

Waddell, L. R. (2014). Using culturally ambitious teaching practices to support urban mathematics teaching and learning. *Journal of Praxis in Multicultural Education*, *8*(2), 2.

Wilson, J., Nazemi, M., Jackson, K., & Wilhelm, A. G. (2019). Investigating Teaching in Conceptually Oriented Mathematics Classrooms Characterized by African American Student Success. *Journal for Research in Mathematics Education*, *50*(4), 362-400.

Yuan, K. H., & Bentler, P. M. (2005). Asymptotic robustness of the normal theory likelihood ratio statistic for two-level covariance structure models. *Journal of Multivariate Analysis,* *94*(2), 328-343.