Setting the Foundation for Equitable Classroom Practices: Enhancing Teacher Comfort and Preparedness for Teaching Diverse Students using the *Math for All* Professional Development Program

Babette Moeller, Education Development Center

Teresa Duncan, Deacon Hill Research Associates

Jason Schoeneberger, ICF

John Hitchcock, Westat

Matthew McLeod, Education Development Center

This paper was developed with funding from two grants from the U.S. Department of Education (grant numbers R305A140488 and U411B180037). The contents of this paper do not necessarily represent the policy or views of the U.S. Department of Education, nor do they imply endorsement by the U.S. Department of Education.

Citation: Moeller, B., Duncan, T., Schoeneberger, J., Hitchcock, J., & McLeod, M. (2022, September). *Setting the foundation for equitable classroom practices: Enhancing teacher comfort and preparedness for teaching diverse students using the Math for All professional development program*. Paper presented at the annual research conference of the National Council of Teachers of Mathematics, Los Angeles, CA.

Setting the Foundation for Equitable Classroom Practices: Enhancing Teacher Comfort and Preparedness for Teaching Diverse Students using the *Math for All* Professional Development Program

Background

Research shows that teacher quality is the single most powerful influence on student learning (e.g., Nye et al., 2004; O'Dwyer et al., 2010; Rivkin et al., 2005). Yet teachers report that they are not well prepared to implement standards-based mathematics education with the heterogeneous groups of students often found in general education classrooms, including students with disabilities and students with different capabilities and needs. In a national survey of science and mathematics teachers, Banilower et al. (2018) found that only 41% of elementary school math teachers felt well prepared to differentiate mathematics instruction to meet the needs of diverse learners, and 15% felt well prepared to incorporate students' cultural backgrounds into mathematics instruction. Notably, these results are consistent with an earlier national survey, suggesting this is a long-term concern (Banilower et al., 2013).

Given teachers' self-reported lack of preparation to teach high-quality mathematics to diverse learners, it is not surprising that mathematics achievement in the United States is low, especially for high-need student populations. According to recent data from the National Assessment of Educational Progress (NAEP) (U.S. Department of Education, 2019), mathematics achievement levels for students with disabilities, and those from low-income families are among the lowest of all U.S. students (Table 1).

Table 1. *2019 NAEP Mathematics Assessment Results*

Student Group Scoring Proficient or Higher	Grade 4	Grade 8
All Students	41%	34%
Low Income (National School Lunch Program)	26%	18%
Students with Disabilities	17%	9%

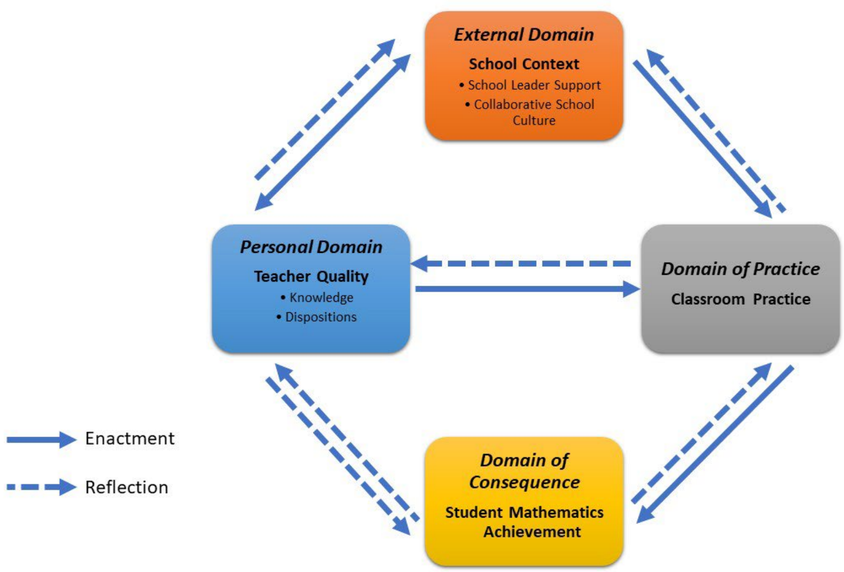
This is alarming. Mathematics is essential for functioning in everyday life and is a prerequisite to many 21st-century careers. In addition, research has shown that mathematics achievement is closely linked with overall student success, such as achievement in high school, high school graduation, college readiness, and students' career aspirations (e.g., Balfanz et al., 2007; Lee, 2012; Shapka et al., 2006; Siegler et al., 2012). Thus, the low mathematics achievement of high-need student populations threatens to limit their opportunities to excel in an increasingly technology-based society.

Theoretical Framework

The goal of this paper is to present data from two randomized controlled trials (RCTs) of the *Math for All* professional development (PD) program. *Math for All* is an intensive PD program designed to help general and special education teachers in Grades K–5 personalize rigorous mathematics instruction for a wide range of learners, including students who are low performing, and students with disabilities. PD programs that help teachers understand and respond to individual students' strengths and needs can support equity in the classroom by building teachers' preparedness and comfort, which enhances the likelihood of teachers' engaging in differentiated instruction to support all students.

As we seek to understand the impacts of PD on teacher and student outcomes, it is important to "open the black box" and flesh out the mechanisms by which PD can affect teacher practice (which in turn, affects student achievement). Clarke & Hollingsworth (2002) propose that teacher quality, along with school context, classroom practice, and student mathematics achievement, are the key components to consider when conceptualizing teachers' professional growth, which is an ongoing, dynamic, and interactive enterprise. As shown in Figure 1, comfort and preparedness are dispositions that along with teacher knowledge, comprise teacher quality (Clarke & Hollingsworth, 2002).

Figure 1. *Dynamic model of teachers' professional growth (adapted from Clark & Hollingsworth 2002 and Goldsmith et al. 2013)*



Intervention

PD programs that are embedded in subject area content and focus on how students learn content have been found more likely to change classroom practices and enhance student outcomes, relative to approaches that focus mainly on the processes for delivery of instruction (Cohen & Hill, 1998; Corcoran, 1995; Garet et al., 2001; Kennedy, 1998). *Math for All* prepares grade K–5 teachers to help students with diverse strengths and needs—including those with disabilities—who are being served in general education classrooms achieve high-quality, standards-based learning outcomes in mathematics. The *Math for All* program is designed to have a direct impact on teachers' knowledge, skills, and classroom practice. The PD introduces teachers to a neurodevelopmental framework (Barringer et al., 2010; Levine, 2002; Pohlman, 2008) as a lens for better understanding individual students' strengths and needs and the demands of mathematical activities. This framework includes eight constructs related to learning processes (i.e., attention, temporal-sequential ordering, spatial-ordering, memory, language, neuromotor function, social cognition, and higher order cognition). Those who use the framework are encouraged to think through how these constructs interact when student learn. It also engages teachers in in-depth analyses of math lessons, including examination of their mathematical goals, and different kinds of instructional strategies and teaching practices that support the attainment of these goals while being attuned to individual students' strengths and needs.

Math for All was developed by Bank Street College and EDC with funding from the National Science Foundation and is published by Corwin Press (Moeller et al., 2011a; 2011b; 2012; 2013a; 2013b). *Math for All* incorporates several components that randomized controlled trials (RCTs) and quasi-experimental studies (QEDs) have shown to be effective for supporting elementary school teachers' professional learning and for improving student achievement, particularly teacher collaboration for instructional planning and peer coaching (cf. Stevens & Slavin, 1995).

Research Questions

The research questions we addressed in our analyses were as follows:

1. What is the impact of *Math for All* on K-5 mathematics teachers' reports of *preparedness* for teaching diverse students?
2. What is the impact of *Math for All* on K-5 mathematics teachers' reports of *comfort* with teaching diverse students?
3. Does the impact of *Math for All* on K-5 mathematics teachers reports of preparedness and comfort vary by *mode of PD delivery* (*Math for All* developers vs. local staff developers)?

4. Does the impact of *Math for All* on K-5 mathematics teachers reports of preparedness and comfort vary by *grade level* (early elementary K-2 vs. upper elementary grades 3-5)?
5. Does the impact of *Math for All* on K-5 mathematics teachers reports of preparedness and comfort vary by *locale*?
6. How are preparedness and comfort related to teachers' reports of instructional practices?

Data Sources

RCT #1. In fall 2014, the Institute of Education Sciences (IES) funded an efficacy trial of *Math for All* to help build the knowledge base on the impact of PD interventions (Author et al., 2018). The study took place during the 2015–16 and 2016–17 school years, involving 32 schools from a large urban school district in the Midwest and 80 4th and 5th grade general and special education teachers. Because sample attrition occurred between the first and second years of the study, we focus on results recorded after the first year of the study, where causal validity is strongest. Note that this RCT did find a compelling school-level effect size on achievement (ES = 0.327) but this was not a statistically significant result because it is a cluster-level contrast with a sample size of 32.

RCT #2. In fall 2018, the authors received a mid-phase Education Innovation Research (EIR) grant from the U.S. Dept. of Education. The overall goal of this project is to implement, test, and refine strategies for regionally expanding *Math for All* in a variety of settings and with diverse populations. We report results from the 2019-20 school year, involving 17 schools (11 in a large urban district and 6 in several rural districts in a midwestern state) and 153 general and special education teachers in grades K through 5. *Math for All* PD was randomly assigned to be delivered either at the K-2 or 3-5 grade band. In other words, all schools had access to the PD; but the PD was offered only within one of two grade band options. Student achievement data are not available for the 2019-20 school year because state assessments were cancelled due to the COVID-19 pandemic.

Measures

For both RCTs, we administered a teacher survey at the beginning and end of the school year that included measures of teacher preparedness and comfort (Table 2). We also administered instructional logs that asked teachers to report on their practices at the end of several targeted weeks (Table 3). The number of logs completed by individual teachers varied, so we computed averages for teacher practices to correlate with preparedness and comfort. Internal consistency of all three scales is high, as measured by the Cronbach alphas shown in Table 4.

Table 2. *Preparedness and Comfort Scales*

a. Teaching standards-based math to students with disabilities.
b. Identifying the math strengths of students with disabilities.
c. Identifying the math needs of students with disabilities.
d. Understanding the mathematics of the lessons I teach.
e. Analyzing the demands of mathematical tasks on students.
f. Determining the goals of the math lessons I teach.
g. Understanding learning trajectories in mathematics (how the math I teach relates to what students learned before and what they will learn later).
h. Selecting specific strategies to address the strengths of students with disabilities in math.
i. Selecting specific strategies to address the needs of students with disabilities in math.
j. Adapting math lessons for students with disabilities to help them meet standards-based goals.
k. Collaborating with my colleagues when planning math lessons.

Note. Items are rated on 1-5 Likert scales, anchored by 1=not at all prepared to 5=very prepared, or 1=not at all comfortable to 5=very comfortable

Table 3. *Instructional Practices Measure*

a. Observe individual students to identify strengths and needs.
b. Analyze student work samples to identify strengths and needs.
c. Share assessment data with students.
d. Encourage students to reflect on their learning.
e. Give information using multiple modalities (e.g., visual, verbal, written).
f. Use graphics and visual organizers to represent math concepts and problems.
g. Allow students to express their ideas in multiple modalities.
h. Offer assignments with different levels of difficulty for different students.
i. Allow for students to engage with the lesson materials in multiple ways.
j. Have students explain their thinking by talking, writing, or drawing the steps they used in solving a problem.
k. Have students solve problems and discuss mathematics in small groups.
l. Group students so they can provide peer support.
m. Clearly communicate expectations for learning.
n. Systematically and explicitly teach the steps and strategies for problem solving.
o. Evaluate how specific strategies are working for individual students.
p. Reflect on my practice.

Note. Items are rated on the following frequency scale: 0=Never; 1=Once; 2=Twice; 3=Three Times; 4=Four Times; 5=Daily

Table 4. *Scale Reliability Statistics*

	Standardized Cronbach Alpha			
Scale	**Fall 2015**	**Spring 2016**	**Fall 2019**	**Spring 2020**
Preparedness	0.925	0.943	0.909	0.910
Comfort	0.930	0.954	0.928	0.926
Instructional Practices	0.841	0.881	0.886	0.946

Results

Across both RCTs, we tested multilevel models (MLMs) that included the baseline scores and an indicator for study condition (*Math for All* PD or business as usual [BAU]). Please refer to the figure notes for details about the statistical analyses.

The *Math for All* PD vs. BAU group differences in comfort and preparedness are presented in Figures 2 and 3.

a) Graphs 2A and 2B (preparedness) and 3A and 3B (comfort) compare the overall results of the two RCTs, where mode of PD delivery differed: the PD was delivered by *Math for All* developers in RCT #1, and by local staff developers in RCT #2 ("train the trainer").

b) Graphs 2C and 2D (preparedness) and 3C and 3D (comfort) break out the results of Graphs 2B and 3B to show differences in results between the lower and upper elementary grade bands.

- Results from the 2015-2016 school year, grades 4 and 5 teachers (n = 80)
- Results from the 2019-20 school year, K-5 teachers (n = 153)
- Results from the 2019-20 school year, K-2 teachers (n = 94)
- Results from the 2019-20 school year, grades 3-5 teachers (n = 59)

c) Graphs 2E and 2F (preparedness) and 3E and 3F (comfort) break out the results of Graphs 2B and 3B to show differences in results between the rural and urban teachers.

- Results from the 2019-20 school year, rural teachers (n = 56)
- Results from the 2019-20 school year, urban teachers (n = 97)

A summary of effect sizes across the subgroups in RCT #2 is shown in Table 5. Correlations between preparedness, comfort, and instructional practices for both studies are shown in Table 6.

RQ #1: Impact of Math for All on K-5 teachers' reports of preparedness to teach mathematics to diverse students. In both studies, we found statistically significant results of the *Math for All* PD on teachers' preparedness to teach mathematics to diverse students. Teachers in the BAU groups reported either lower or unchanged levels of preparedness to teach mathematics to diverse students. In contrast, teachers who participated in the *Math for All* PD reported substantive increases in their preparedness (Figures 2A through 2F).

RQ #2: Impact of Math for All on K-5 teachers' reports of comfort with teaching mathematics to diverse students. In both studies, we found statistically significant results of the *Math for All* PD on teachers' comfort with teaching mathematics to diverse students. Teachers in the BAU groups reported either lower or unchanged levels of comfort, while teachers who participated in the *Math for All* PD reported significant increases in their comfort with teaching mathematics to diverse students (Figures 3A through 3F).

RQ #3: Impact of Math for All on K-5 teachers' preparedness and comfort, by mode of PD delivery. As noted above, the PD was delivered by *Math for All* developers in RCT #1, and by local staff developers in RCT #2 (again, the "train the trainer" approach). In light of the results presented under RQ #1 and RQ #2, the effect of *Math for All* was stronger when the PD was delivered by *Math for All* developers than when delivered by local staff developers. Effect sizes were 0.541 vs. 0.380 for preparedness and 0.666 vs. 0.285 for comfort.

RQ #4: Impact of Math for All on K-5 teachers' preparedness and comfort, by grade level. When results for RCT #2 (Graphs 2B and 3B) were split by lower and upper grade levels (K-2, Graphs 2C and 3C; and grades 3-5, Graphs 2D and 3D), the effect of *Math for All* was stronger among the grades 3-5 teachers than among the K-2 teachers. Effect sizes were 0.488 vs. 0.322 for preparedness and 0.300 vs. 0.241 for comfort.

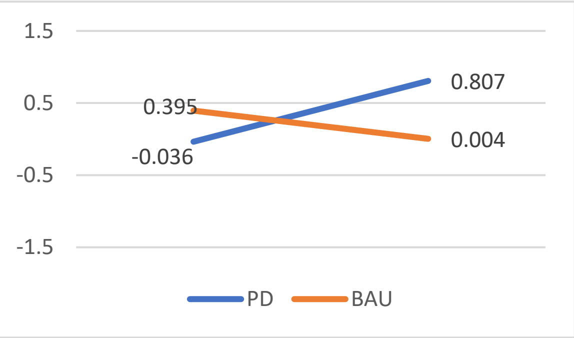
RQ #5: Impact of Math for All on K-5 teachers' preparedness and comfort, by locale. When results for RCT #2 (Graphs 2B and 3B) were split by lower and upper grade levels (K-2, Graphs 2E and 3E; and grades 3-5, Graphs 2F and 3F), the effect of *Math for All* was stronger among the urban teachers than among the rural teachers. Effect sizes were 0.442 vs. 0.162 for preparedness and 0.329 vs. 0.046 for comfort).

RQ #6: Relationships between Teacher Dispositions and Instructional Practices. Correlations between teacher instructional practices with comfort and preparedness to teach diverse students are displayed in Table 6. Although the correlations between preparedness and comfort and instructional practices are modest, the correlations are consistent across the two studies. In both studies and across sub-groups in RCT #2, the correlations obtained in the spring are stronger than those obtained in the fall. During spring 2020 when the COVID-19 pandemic began, the correlations between preparedness and instructional practices, and between comfort and instructional practices were markedly lower in the rural sample than in the urban sample.

Figure 2. Impact of *Math for All* PD on participating teachers' **preparedness** for teaching mathematics to diverse students

2A. 2015-2016 School Year: Preparedness Results from RCT #1 (Grade 4 and 5 Teachers)

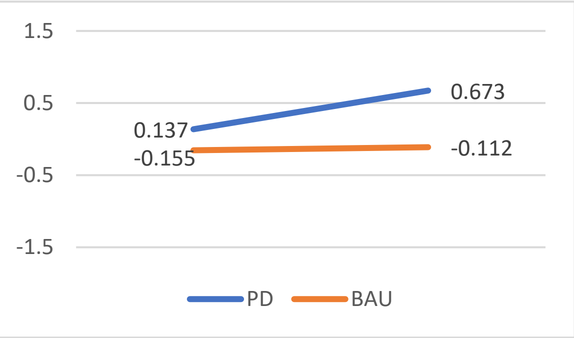
Group	Fall 2015 Preparedness Mean	Fall 2015 Preparedness SD	Spring 2016 Preparedness Unadj. Mean	Spring 2016 Preparedness Adj. Mean	Spring 2016 Preparedness SD	Hedges g	p-value
PD	-.036	1.193	.748	.807	1.322		
BAU	.395	1.180	.060	.004	1.623	.541	.013



Note. Multilevel model with the Preparedness dependent variable, including the treatment indicator, the baseline Preparedness variable, enjoy teaching math, an indicator denoting the teacher as a SWD teacher, and a school-level mean baseline Preparedness variable. The model also included a random intercept, making use of n =37 BAU teachers and n =43 PD teachers. The baseline difference for the analytic sample was computed as g = -0.36. Because of the size of this baseline difference, this finding should be interpreted with caution.

2B. 2019-2020 School Year: Preparedness from RCT #2 (Grades K – 5 Teachers)

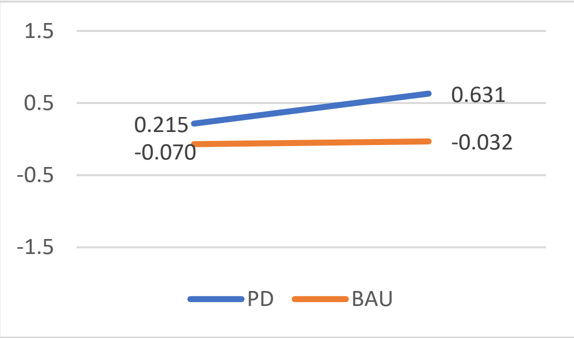
Group	Fall 2019 Preparedness Mean	Fall 2019 Preparedness SD	Spring 2020 Preparedness Unadj. Mean	Spring 2020 Preparedness Adj. Mean	Spring 2020 Preparedness SD	Hedges g	p-value
PD	.137	1.995	.772	.673	2.315		
BAU	-.155	1.617	-.198	-.112	1.530	.380	.007



Note. Multilevel model with cohort-by-district block included in the model as a predictor (n = 58 BAU teachers and n = 95 PD teachers). The baseline difference for the analytic sample was computed as g = 0.156.

2C. 2019-2020 School Year: Preparedness from RCT #2 (Grades K – 2 Teachers)

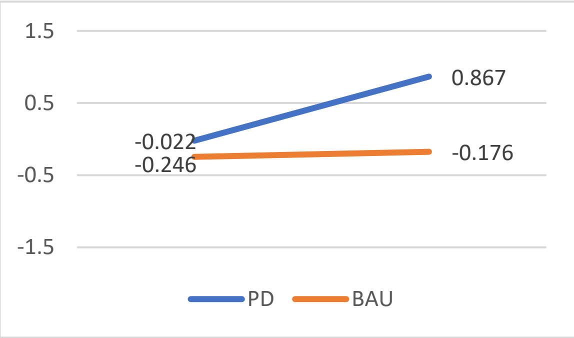
Group	Fall 2019 Preparedness Mean	Fall 2019 Preparedness SD	Spring 2020 Preparedness Unadj. Mean	Spring 2020 Preparedness Adj. Mean	Spring 2020 Preparedness SD	Hedges g	p-value
PD	.215	1.973	.735	.631	2.254		
BAU	-.070	1.548	-.108	-.032	1.474	.322	.074



Note. Multilevel model with cohort-by-district block included in the model as a predictor (*n* = 30 BAU teachers and *n* = 64 PD teachers in the K-2 grade band). The baseline difference for the analytic sample was computed as *g* = 0.153.

2D. 2019-2020 School Year: Preparedness from RCT #2 (Grades 3 – 5 Teachers)

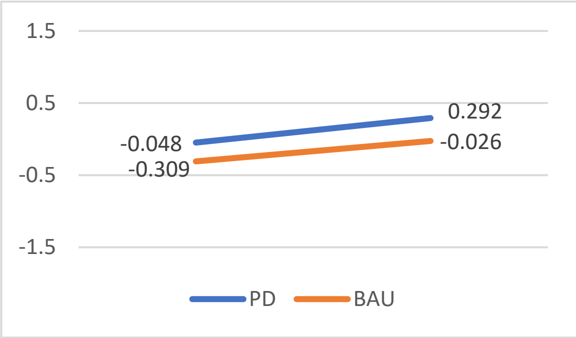
Group	Fall 2019 Preparedness Mean	Fall 2019 Preparedness SD	Spring 2020 Preparedness Unadj. Mean	Spring 2020 Preparedness Adj. Mean	Spring 2020 Preparedness SD	Hedges g	p-value
PD	-.022	2.063	.849	.867	2.472		
BAU	-.246	1.712	-.294	-.176	1.610	.488	.032



Note. Multilevel model with cohort-by-district block included in the model as a predictor (*n* = 28 BAU teachers and *n* = 31 PD teachers in the 3-5 grade band). The baseline difference for the analytic sample was computed as *g* = 0.116.

2E. 2019-2020 School Year: Preparedness from RCT #2 (Rural Teachers)

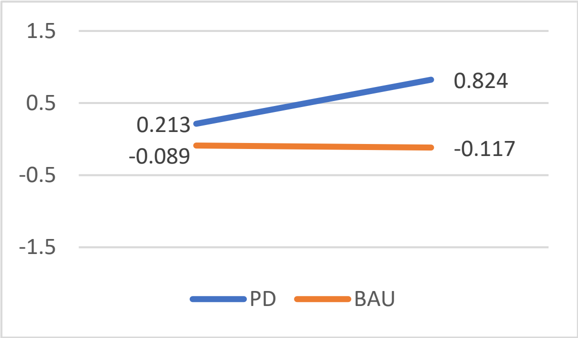
Group	Fall 2019 Preparedness Mean	Fall 2019 Preparedness SD	Spring 2020 Preparedness Unadj. Mean	Spring 2020 Preparedness Adj. Mean	Spring 2020 Preparedness SD	Hedges g	p-value
PD	-.048	2.200	.342	.292	2.288		
BAU	-.309	1.253	-.125	-.026	1.204	.162	.481



Note. Multilevel model with baseline preparedness included in the model as a predictor (*n* = 34 BAU teachers and *n* = 22 PD teachers in the sample). The baseline difference for the analytic sample was computed as *g* = 0.153.

2F. 2019-2020 School Year: Preparedness from RCT #2 (Urban Teachers)

Group	Fall 2019 Preparedness Mean	Fall 2019 Preparedness SD	Spring 2020 Preparedness Unadj. Mean	Spring 2020 Preparedness Adj. Mean	Spring 2020 Preparedness SD	Hedges g	p-value
PD	.213	1.902	1.012	.824	2.314		
BAU	-.089	1.839	-.242	-.117	1.714	.442	.024

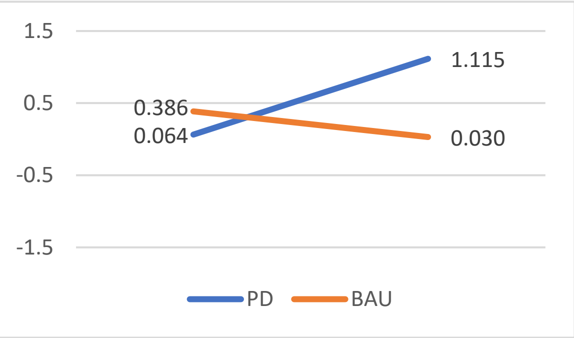


Note. School fixed-effect single-level regression with baseline preparedness included in the model as a predictor (*n* = 61 BAU teachers and *n* = 36 PD teachers in the sample). The baseline difference for the analytic sample was computed as *g* = 0.161.

Figure 3. Impact of *Math for All* PD on participating teachers' **comfort** with teaching mathematics to diverse students

3A. 2015-2016 School Year: Comfort Results from RCT #1 (Grade 4 and 5 Teachers)

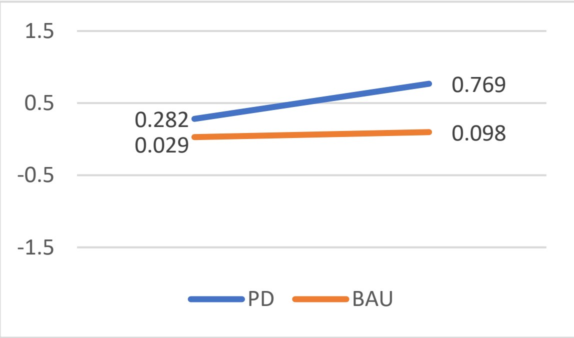
Group	Fall 2015 Comfort Mean	Fall 2015 Comfort SD	Spring 2016 Comfort Unadj. Mean	Spring 2016 Comfort Adj. Mean	Spring 2016 Comfort SD	Hedges g	p-value
PD	.064	1.297	1.088	1.115	1.442		
BAU	.386	1.235	.051	.030	1.800	.666	.005



Note. Multilevel model with the Comfort dependent variable, including the treatment indicator, the baseline Comfort variable, and a variable representing enjoyment for teaching math. The model also included a random intercept, making use of n = 36 BAU teachers and n = 43 PD teachers. The baseline difference for the analytic sample was computed as g = -0.251.

3B. 2019-2020 School Year: Comfort from RCT #2 (Grades K – 5 Teachers)

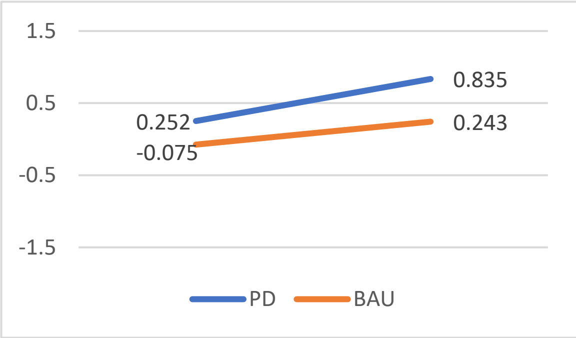
Group	Fall 2019 Comfort Mean	Fall 2019 Comfort SD	Spring 2020 Comfort Unadj. Mean	Spring 2020 Comfort Adj. Mean	Spring 2020 Comfort SD	Hedges g	p-value
PD	.282	2.322	.932	.769	2.480		
BAU	.029	2.132	.115	.098	2.092	.285	.051



Note. Multilevel model with cohort-by-district block included in the model as a predictor (n = 57 BAU teachers and n = 95 PD teachers). The baseline difference was computed as g = 0.112.

3C. 2019-2020 School Year: Comfort from RCT #2 (Grades K – 2 Teachers)

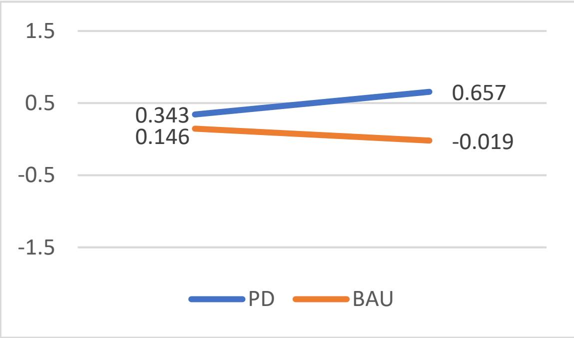
Group	Fall 2019 Comfort Mean	Fall 2019 Comfort SD	Spring 2020 Comfort Unadj. Mean	Spring 2020 Comfort Adj. Mean	Spring 2020 Comfort SD	Hedges g	p-value
PD	.252	2.307	1.005	.835	2.581		
BAU	-.075	1.847	.226	.243	2.077	.241	.182



Note. Multilevel model with cohort-by-district block included in the model as a predictor (n = 30 BAU teachers and n = 64 PD teachers in the K-2 grade band). The baseline difference was computed as g = 0.149.

3D. 2019-2020 School Year: Comfort from RCT #2 (Grades 3 – 5 Teachers)

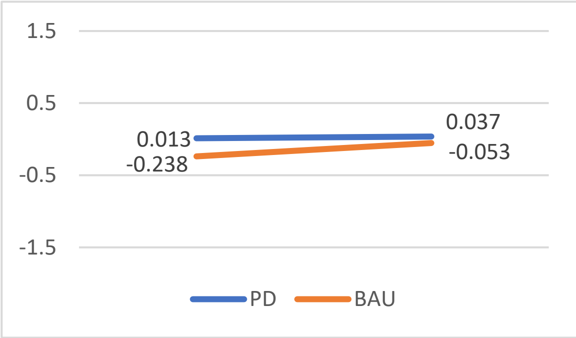
Group	Fall 2019 Comfort Mean	Fall 2019 Comfort SD	Spring 2020 Comfort Unadj. Mean	Spring 2020 Comfort Adj. Mean	Spring 2020 Comfort SD	Hedges g	p-value
PD	.343	2.390	.781	.657	2.289		
BAU	.146	2.441	-.007	-.019	2.142	.300	.224



Note. Multilevel model with cohort-by-district block included in the model as a predictor (n = 27 BAU teachers and n = 31 PD teachers in the 3-5 grade band). The baseline difference was computed as g = 0.081.

3E. 2019-2020 School Year: Comfort from RCT #2 (Rural Teachers)

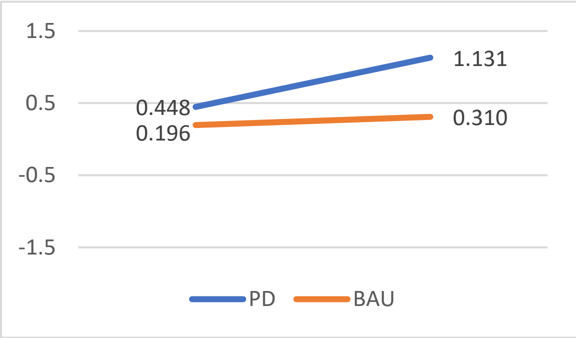
Group	Fall 2019 Comfort Mean	Fall 2019 Comfort SD	Spring 2020 Comfort Unadj. Mean	Spring 2020 Comfort Adj. Mean	Spring 2020 Comfort SD	Hedges g	p-value
PD	-.013	2.484	.021	.037	2.201		
BAU	-.238	1.668	-.142	-.053	1.388	.046	.855



Note. Multilevel model with baseline comfort included in the model as a predictor (*n* = 34 BAU teachers and *n* = 22 PD teachers in the sample). The baseline difference for the analytic sample was computed as *g* = 0.110.

3F. 2019-2020 School Year: Comfort from RCT #2 (Urban Teachers)

Group	Fall 2019 Comfort Mean	Fall 2019 Comfort SD	Spring 2020 Comfort Unadj. Mean	Spring 2020 Comfort Adj. Mean	Spring 2020 Comfort SD	Hedges g	p-value
PD	.448	2.270	1.439	1.131	2.497		
BAU	.196	2.424	.278	.310	2.439	.329	.108



Note. School fixed-effect single-level regression with baseline preparedness included in the model as a predictor (*n* = 61 BAU teachers and *n* = 35 PD teachers in the sample). The baseline difference for the analytic sample was computed as *g* = 0.106.

Table 5. *Summary of Math for All PD v. BAU effect sizes within subgroups*

	Effect Size (Hedges g)	p-value
RCT #1: Preparedness (Graph 2A); PD delivered by PD developers	.541	.013
RCT #2 Preparedness (Graph 2B); PD delivered by local staff developers	.380	.007
RCT #2: Preparedness (Graph 2C): Grades K-2 teachers	.322	.074
RCT #2: Preparedness (Graph 2D): Grades 3-5 teachers	.488	.032
RCT #2: Preparedness (Graph 2E): Rural teachers	.162	.481
RCT #2: Preparedness (Graph 2F): Urban teachers	.442	.024
RCT #1: Comfort (Graph 3A); PD delivered by PD developers	.666	.005
RCT #2 Comfort (Graph 3B); PD delivered by local staff developers	.285	.051
RCT #2: Comfort (Graph 3C): Grades K-2 teachers	.241	.182
RCT #2: Comfort (Graph 3D): Grades 3-5 teachers	.300	.224
RCT #2: Comfort (Graph 3E): Rural teachers	.046	.855
RCT #2: Comfort (Graph 3F): Urban teachers	.329	.108

Table 6. *Correlations between Preparedness, Comfort, and Instructional Practices*

Scale	Preparedness	Comfort	Instructional Practices
RCT #1: 2015-2016 data (grades 4 and 5 teachers)			
Preparedness	—	0.93*	**0.20**
Comfort	0.92*	—	**0.31***
Instructional Practices	**0.11**	**0.07**	—
RCT #2: 2019-2020 data (grades K-5 teachers)			
Preparedness	—	0.86*	**0.32***
Comfort	0.80*	—	**0.38***
Instructional Practices	**0.25***	**0.23***	—
RCT #2: 2019-2020 data (grades K-2 teachers)			
Preparedness	—	0.88*	**0.31***
Comfort	0.83*	—	**0.39***
Instructional Practices	**0.26***	**0.21***	—
RCT #2: 2019-2020 data (grades 3-5 teachers)			
Preparedness	—	0.83*	**0.34***
Comfort	0.77*	—	**0.38***
Instructional Practices	**0.24***	**0.26***	—
RCT #2: 2019-2020 data (rural teachers)			
Preparedness	—	0.89*	**0.11**
Comfort	0.92*	—	**-0.01**
Instructional Practices	**0.24**	**0.13**	—
RCT #2: 2019-2020 data (urban teachers)			
Preparedness	—	0.77*	**0.25***
Comfort	0.84*	—	**0.22***
Instructional Practices	**0.20**	**0.29***	—

Note. Values below the diagonal are among the fall measures; values above the diagonal are among the spring measures. Asterisks indicate correlations that are statistically significant a $p \leq 0.05$. Readers might wish to interpret these correlations by squaring them (e.g., 0.31 = .0961, or about 10% of the variance in one measure is accounted for by variance in the other).

Conclusion and Significance

The body of work presented here provides strong causal evidence based on two RCTs that *Math for All* had an impact (i.e., statistically significant, medium to large effect sizes) on teachers' self-reported sense of comfort and preparedness for teaching diverse students. This is critical given repeated survey findings that most teachers report feeling underprepared to work with diverse students (Banilower et al., 2013; 2018) and NAEP findings show large numbers of students are underperforming in mathematics. As argued earlier, poor mathematics achievement can impact long-term individual student outcomes and national competitiveness.

The consistency of results across both RCTs indicates stability in the patterns of findings, which point to positive impacts of the *Math for All* program on elementary teachers' preparedness for and comfort in teaching mathematics to diverse learners. We do note that the *magnitude* of the effect does vary across subsamples, where impacts of *Math for All* are stronger among grades 3-5 teachers and teachers in urban schools.

These analyses also uncovered consistent, modest, positive correlations between preparedness, comfort, and instructional practices. The positive correlations between teacher dispositions and instructional practices complement prior reports of the positive impact of *Math for All* on instructional practices, as measured via classroom observations of mathematics lessons and scores using the CLASS rubric (Author et al., 2018). The fact that the results were similar across the two RCTs even though the modes of delivery differed (PD led by developers in RCT #1; by local staff developers and teacher leaders in RCT #2) also speaks to the scalability of the *Math for All* program.

Two key limitations are worth noting. The first is that in RCT #1 there are large baseline differences between the *Math for All* and BAU groups in the comfort and preparedness contrasts. The second limitation is that we do not present student-level achievement outcomes (recall however that the first RCT did show a compelling student-level effect although it was not statistically significant). We were unable to collect student-level data for the second RCT because of COVID-19, but we do have funding necessary under RCT #2 to carry out research with another cohort of schools and teachers.

We should also keep in mind that in spring 2020, schools were closing and/or pivoting to emergency remote instruction because of the COVID-19 pandemic. We witnessed how our participating rural schools were particularly hard-hit by the closures and challenges of remote instruction, so the weaker effects of the *Math for All* program on preparedness and comfort, and the lower correlations between preparedness, comfort, and instructional practices should be interpreted within this context.

As stated previously, in order to understand the impacts of PD on teacher and student outcomes, we believe it is important to "open the black box" and flesh out the mechanisms by which PD can affect teacher practice (which in turn, affects student achievement). The data presented here suggest that teacher preparedness and comfort may be key mediators to consider in our models of teacher quality (cf. Clarke & Hollingsworth, 2002).

References

Balfanz, R., Herzog, L., & Mac Iver, D. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist*, *42*(4), 223–235. https://doi.org/10.1080/00461520701621079

Banilower, E. R., Smith, P. S., Malzahn, K. A., Plumley, C. L., Gordon, E. M., & Hayes, M. L. (2018). *Report of the 2018 National Survey of Science and Mathematics Education (NSSME+)*. Retrieved from Chapel Hill, NC: http://horizon-research.com/NSSME/.

Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). *Report of the 2012 National Survey of Science and Mathematics Education.* Horizon Research, Inc. http://www.horizon-research.com/2012nssme/wp-content/uploads/2013/02/2012-NSSME-Full-Report1.pdf

Barringer, M-D., Pohlman, C., & Robinson, M. (2010). *Schools for all kinds of minds: Boosting student success by embracing learning variation*. John Wiley & Sons, Inc.

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, *18*, 947-967. https://doi.org/10.1016/S0742-051X(02)00053-7

Cohen, D. K. & Hill, H. C. (1998). *Instructional policy and classroom performance: The mathematics reform in California* (CPRE RR-39). Consortium for Policy in Education, University of Pennsylvania.

Corcoran, T. B. (1995). *Helping teachers teach well: Transforming professional development*. Consortium for Policy Research in Education RB-16. Rutgers, State University of New Jersey.

Duncan, T.G., Moeller, B., Schoeneberger, J., & Hitchcock, J.H. (2018). *Assessing the impact of the Math for All professional development program on elementary school teachers and their students*. Available from: https://bit.ly/2sqzUer

Garet, M. S., Porter, A. C., Desimone, L., Birman, B., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*(4), 915–945. https://doi.org/10.3102/00028312038004915

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*, 5-36. https://doi.org/10.1007/s10857-013-9245-4

Kennedy, M. (1998). *Form and substance of in-service teacher education* (Research Monograph No. 13). National Institute for Science Education, University of Wisconsin-Madison.

Lee, J. (2012). College for all: Gaps between desirable and actual P–12 math achievement trajectories for college readiness. *Educational Researcher*, *41*(2), 43–55. https://doi.org/10.3102/0013189X11432746

Levine, M. (2002). *A mind at a time*. New York, NY: Simon & Schuster.

Moeller, B., Brodesky, A., & Goldsmith, L. (2011b). *Supporting math teacher educators' implementation of curriculum-based professional development programs*. Session conducted at the Annual Meeting of the Association of Mathematics Teacher Educators, January 27–29, 2011, Irvine, CA.

Moeller, B., Dubitsky, B., Cohen, M., Marschke-Tobier, K., Melnick, H., & Metnetsky, L. (2011a). *Mathematics for All: Participant book for grades 3–5*. Corwin Press.

Moeller, B., Dubitsky, B., Cohen, M., Marschke-Tobier, K., Melnick, H., & Metnetsky, L. (2012). *Mathematics for All: Facilitator guide for grades 3–5*. Corwin Press.

Moeller, B., Dubitsky, B., Cohen, M., Marschke-Tobier, K., Melnick, H., & Metnetsky, L. (2013a). *Mathematics for All: Facilitator guide for grades K–2*. Corwin Press.

Moeller, B., Dubitsky, B., Cohen, M., Marschke-Tobier, K., Melnick, H., & Metnetsky, L. (2013b). *Mathematics for All: Participant book for grades K–2*. Corwin Press.

Nye, B., Konstantopoulos, S., & Hedges. L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*(3), 237–257. https://doi.org/10.3102/01623737026003237

O'Dwyer, L. M., Masters, J., Dash, S., DeKramer, R. M., Humez, A., & Russell, M. (2010). *e-Learning for educators: Effects of on-line professional development on teachers and their students: Findings from four randomized trials*. Boston College, Technology and Assessment Study Collaborative. Retrieved from http://www.bc.edu/research/intasc/PDF/EFE_Findings2010_Report.pdf

Pohlman, C. (2008). *Revealing minds: Assessing to understand and support struggling learners*. Jossey-Bass.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458. https://doi.org/10.1111/j.1468-0262.2005.00584.x

Shapka, J. D., Domene, J. F., & Keating, D. P. (2006). Trajectories of career aspirations through adolescence and young adulthood: Early math achievement as a critical filter. *Educational Research and Evaluation*, *12*(4), 347–358. https://doi.org/10.1080/13803610600765752

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M.…Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological Science 23*(7), 691–697. https://doi.org/10.1177/0956797612440101

Stevens, R. J., & Slavin, R. E. (1995). The cooperative elementary school: Effects on students' achievement, attitudes, and social relations. *American Educational Research Journal*, *32*(2), 321–351. https://doi.org/10.3102/00028312032002321

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (NCES). (2019). *National Assessment of Educational Progress (NAEP), 2019 mathematics assessment data, 4th grade and 8th grade*. Retrieved from https://www.nationsreportcard.gov/highlights/mathematics/2019/