Assessing the impact of the Math for All professional development program on elementary school teachers and their students

Teresa Duncan

Deacon Hill Research Associates

Babette Moeller

Education Development Center

Jason Schoeneberger

ICF

John Hitchcock

Abt Associates

This research was generously supported by a grant from the U.S. Dept. of Education, Institute of Education Sciences (Grant No. R305A140488 awarded to Education Development Center). Opinions expressed in this paper are the authors' and not of the U.S. Dept. of Education. Please address all correspondence to Babette Moeller (bmoeller@edc.org).

Table of Contents

Exhibits

Appendix

Abstract

Math for All (MFA) is an intensive professional development (PD) program, consisting of five one-day workshops and classroom-based assignments, providing a total of 50 hours of PD, typically conducted over the course of one school year. The program shows teams of general and special education teachers how to collaboratively plan and adapt math lessons to help all students achieve high-quality, standards-based learning outcomes in mathematics.

In fall 2014, the Institute of Education Sciences (IES) funded a randomized controlled trial (RCT) trial of MFA to help build the knowledge base on the impact of PD interventions. The study took place during the 2015–2016 and 2016–2017 academic years, involving 32 Chicago Public Schools, 96 fourth- and fifth-grade general and special education teachers, and approximately 1,500 fourth- and fifth- grade students. Because of attrition that occurred between first and second years of the study, this brief focuses on results recorded after the first year of the study, where causal validity is strongest.

Our team examined the impact of MFA on teacher outcomes (i.e., knowledge, beliefs, and classroom practice) and student achievement in mathematics (as measured by Northwest Education Association [NWEA] scale scores). Effect sizes range from 0.11 to 0.98.

Although we found an effect favoring the MFA group on **teacher mathematical content knowledge**, we cannot conclude from the pattern observed that MFA improves teacher mathematical content knowledge. We did find statistically significant, positive effects of MFA on teachers' reports of **preparedness and comfort in teaching diverse students** (including students with disabilities). Although MFA teachers scored higher on **emotionally supportive classroom practices**, **instructional support, classroom organization, and student engagement**, the classroom observation data were underpowered and these findings did not reach statistical significance.

Grades 4 and 5 students' **mathematics achievement** were examined at the school and individual student levels. The school-level (or aggregated) analysis assessed MFA's impact on student achievement on **all** grade 4 and grade 5 students at the 32 study schools. The effect size was 0.33, but not statistically significant, likely because of the small sample size of 32 schools. Nevertheless, we are encouraged by this finding, because the results favor the MFA group, even in a cluster analysis that dilutes the treatment effect by including students of nonparticipating grade 4 and grade 5 teachers at the treatment schools. The student-level results also favor the treatment group, but the differences were not statistically significant. When grade level was examined as a moderator, we found different patterns between the grade 4 and the grade 5 samples. In **grade 4**, students whose teachers participated in the MFA PD had significantly higher posttest NWEA Measures of Academic Progress scores than students whose teachers were in the business-as-usual group. In **grade 5**, there were very small mean differences between the MFA and comparison groups, which were not statistically significant.

The large impacts on teacher dispositions and the grade-level interaction on student achievement that we observed in our data indicate that teacher mediators and contextual factors may merit greater attention in PD theories of change. The results suggest that the pathway from PD to teacher practice may not necessarily be a linear progression; perhaps a more dynamic model can capture the relationships more accurately.

Introduction

Professional development (PD) that is embedded in subject area content and how students learn that content has been found more likely to be related to changes in classroom practices and enhanced student outcomes than traditional approaches that focus mainly on the processes for delivery of instruction (Cohen & Hill, 1998; Corcoran, 1995; Garet, Porter, Desimone, Birman, & Yoon, 2001; Kennedy, 1998). A small number of PD programs—in particular, Math for All (MFA)—do integrate learning about how to improve instruction for the wide range of students in inclusion classrooms in the context of mathematics subject matter (e.g., Brodesky, Gross, McTigue, & Palmer, 2007; Moeller et al., 2012). However, there is a paucity of rigorous studies that link PD to student outcomes (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007).

In fall 2014, the Institute of Education Sciences (IES) funded an efficacy trial of MFA to help build the knowledge base on the impact of PD interventions. A small pilot of 20 teachers and 339 students in four Chicago Public Schools (CPS) was conducted between January and June 2015. The full study took place during 2015–2016 (implementation year) and 2016–2017 (impact year), involving 32 CPS schools, 96 fourth- and fifth-grade general and special education teachers, and approximately 1,500 fourth- and fifth-grade students (details are provided in Exhibits 1–5).

In this randomized controlled trial (RCT), our research team examined the impact of MFA on teacher outcomes (i.e., knowledge, beliefs, and classroom practice) and student achievement in mathematics (as measured by Northwest Education Association [NWEA] scale scores). The purpose of this brief is to report findings from the first year of the full study[1] and address the following Year 1 (implementation year) research questions:

> Research Question #1. Does participation in MFA PD, compared to business-as-usual (BAU) experiences of a control group, result in greater teacher **mathematical content knowledge** after the completion of the PD?
>
> Research Question #2. Does participation in MFA PD, compared to BAU experiences of a control group, improve teachers' **comfort and preparedness** to teach mathematics to diverse students (including those with disabilities) after the completion of the PD?
>
> Research Question #3. Does participation in MFA PD, compared to the BAU experiences of a control group, result in improved **mathematics classroom practice** after the completion of the PD?
>
> Research Question #4. Does the use of an MFA approach in the classroom result in improved **student achievement in mathematics after one year of intervention exposure**?

[1] Casual inferences from Year 2 findings will be harder to establish because of attrition that occurred between the first and second years of the study. CPS experienced an unanticipated, serious, and long-term labor dispute that increased in severity during the summer of 2015 and was not even partially resolved until the fall of 2016. This dispute led to considerable administrator turnover, which led to a loss of several principals who had agreed to participate in the study (i.e., for the full two years, as well as to facilitate teacher participation and to accept random assignment into study conditions). Study schools also experienced teacher loss and associated changes in school and grade composition at the student level (more details are offered in the Methods section). Year 2 findings will therefore not meet What Works Clearinghouse (WWC) standards, although some of the outcome analyses might meet standards with reservations. Note that teacher-level outcomes at the end of the first year are considered confirmatory, as they represent an opportunity to observe immediate PD impacts. Confirmatory student-level outcomes were meant to focus on outcomes at the end of Year 2, focusing on those students who had two years of study exposure.

The Math for All (MFA) Professional Development Program

MFA is an intensive PD program, consisting of five one-day workshops and classroom-based assignments, providing a total of 50 hours of PD, typically conducted over the course of one school year. The program shows teams of general and special education teachers how to collaboratively plan and adapt math lessons to help all students achieve high-quality, standards-based learning outcomes in mathematics. Although the intervention was designed with a focus on improving mathematics education for students with disabilities, all students are thought to benefit from instruction individualized to their specific learning strengths and needs.

MFA was designed based on a best-practices model of PD (e.g., Darling-Hammond, Hyler, & Gardner, 2017; Loucks-Horsley, Hewson, Love, & Stiles 2010; Garet et al., 2001). In addition, it incorporates components that rigorously designed research has shown to be effective for supporting elementary school teachers' professional learning and improving student achievement, including:

- Lesson study (e.g., Lewis & Perry, 2017)
- Teacher collaboration for instructional planning and peer coaching (e.g., Stevens & Slavin, 1995)
- Videocase-based analysis of practice (e.g., Taylor, Roth, Wilson, Stuhlsatz, & Tipton, 2017)
- Ongoing formative assessment of students (e.g., Espin, Shin, & Busch, 2000; Fuchs & Fuchs, 1986; Fuchs & Fuchs, 2002)
- Extended duration (e.g., Yoon et al., 2007).

MFA differs from other commonly used approaches to PD in several important ways:

- Rather than focusing only on students with disabilities, MFA is designed to help enhance teachers' preparation to better reach all students, including students with and without disabilities. The underlying assumption is that students with disabilities are not fundamentally different from those without disabilities. Helping teachers to better understand the strengths and needs of individual students, and to differentiate instruction based on deep understanding of mathematical goals and different students' strengths and needs, is expected to benefit students with and without disabilities alike.
- MFA is designed for both general and special education teachers, and the collaboration between teams of special and general education teachers is an integral part of the PD. This contrasts with approaches that target general education and special education teachers separately, typically with general education teachers receiving PD in content areas and special education teachers in the delivery of instructional strategies (Birman et al., 2007).
- MFA deeply integrates learning about differentiating instruction into the context of specific, standards-based mathematics content. MFA focuses on enhancing teachers' preparation to make decisions about how to adapt math lessons based on careful consideration of individual students' strengths and needs and the demands of the mathematical activities, while also maintaining the standards-based learning goals of the lesson. This contrasts with other approaches, such as PD in differentiated instruction, that focus on the delivery of instructional strategies across the curriculum.
- MFA is more comprehensive and intensive than the PD that teachers typically participate in to learn how to better meet the needs of students with disabilities. On average, teachers spend only

3.4 hours on this topic, typically in a single session (Birman et al., 2007). MFA developers deliver 50 hours of PD over the course of a school year.

- MFA is not focused on the delivery of a specific curriculum. Instead, MFA uses standards-based case lessons that were selected from various K–5 math curricula to engage teachers in reflection on how to make standards-based mathematics content accessible to diverse learners in different contexts. MFA also introduces teachers to a process of collaborative lesson planning that they apply, as part of the PD, to the particular standards-based curriculum they are using in their school district.

MFA is intended to be co-facilitated by two staff developers, one whose expertise is in mathematics and the other in special education. The PD is facilitated either by the developers or by district-based staff developers who utilize published PD materials (Moeller et al., 2011, 2012, 2013). The materials that support facilitators in the implementation of the MFA program were published by Corwin Press in 2012 and 2013.

Theory of Change

A diagram illustrating the logic model that informed the design of this study and that links key features of the intervention to outcomes is included in **Appendix A**. The MFA program is designed to have a direct impact on teachers' knowledge, skills, and classroom practice. The PD introduces teachers to a neurodevelopmental framework (Barringer, Pohlman, & Robinson, 2010; Levine, 2002; Pohlman, 2008) as a lens for better understanding individual students' strengths and needs and the demands of mathematical activities. 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.

Teacher outcomes for MFA teachers include (1) being more knowledgeable about mathematical content for teaching individual students' neurodevelopmental strengths and needs, and instructional strategies that help to make high-quality mathematics content more accessible for individual students; (2) feeling more prepared to teach students with disabilities; and (3) feeling more comfortable teaching students with disabilities. Improved knowledge, preparedness, and comfort about students' neurodevelopmental strengths and needs, and the goals and demands of mathematical lessons and activities, will help lay the groundwork for teachers to make more informed decisions about how to adapt mathematics lessons to improve their accessibility for a broad range of learners without compromising the rigor of the standards-based goals.

MFA is also designed to have an impact on teaching practice. Classroom-based assignments introduce teachers to a process of collaborative lesson planning that involves teams of general and special education teachers in observing individual students, and in planning, implementing, and reflecting on adaptations for specific mathematics lessons. We expect that these activities will have an impact on teaching practice because they allow teachers to apply in their own classrooms the knowledge and skills they have learned in PD sessions. Active, practice-based approaches to PD have been found to be more responsive to teachers' needs and goals and to how they learn (Ball, 1996; Brown, Bransford, & Cocking, 1999), and more likely to influence changes in teaching practices (e.g., Darling-Hammond, 1997; Loucks-Horsley, Hewson, Love, & Stiles, 1998). Similarly, PD that is intense and sustained over the course of a school year has been found to have a greater chance of transforming teaching practices and student learning than does the traditional approach of offering isolated PD workshops (e.g., Cohen & Hill, 2001; Desimone et al. 2002; Wei et al., 2009). Teacher collaboration and collective participation in PD also contribute to improved classroom practice, as joint work provides opportunities for teachers to share expertise and to engage in a process of inquiry and reflection into practice (e.g., Wei, Darling-Hammond, & Adamson, 2010).

Key outcomes of the MFA PD for teachers' classroom practices include (1) the ongoing assessment of individual students; (2) adapting mathematics lessons to build on students' strengths and weaknesses while maintaining the rigor of the mathematics content; (3) the use of instructional strategies, classroom structures, and materials that are responsive to individual students' strengths and needs; (4) supportive teacher-student interactions; and (5) ongoing collaborative lesson planning between general and special education teachers.

Improved student achievement in mathematics is a key outcome for MFA. We expect that improved student achievement will be a result of teachers' enhanced understanding of individual students' neurodevelopmental strengths and needs in mathematics, and of improved instruction that is responsive to how individual students learn and that makes standards-based mathematics lessons accessible to a broad range of learners.

Methods

Sample and Design

This study was designed as a two-year cluster RCT with schools randomized into study conditions. School-level assignment was appropriate, given MFA's focus on teacher collaboration under coordinated instructional leadership. The first year of the study (2015–2016) was intended to be the implementation year, during which the PD was delivered and impact on teachers assessed. The second year of the study (2016–2017) was intended to be the impact year, when impact on students would be determined. The idea was that the second year would offer the clearest opportunity to observe student impacts because (1) MFA teachers would have an additional year of experience in using what they learned during the Year 1 PD, and (2) students who had MFA teachers in grade 4 (during Year 1) and in grade 5 (during Year 2) would be most likely to exhibit the benefits of their teachers' enhanced instruction.

Schools. In October 2015, 32 schools in Chicago were randomized into the MFA PD treatment condition or BAU control condition. Two BAU schools and one MFA school left the study in late fall 2015. At the end of spring 2016, another BAU school and two MFA schools were lost, resulting in 13 schools in each condition during Year 2. The Consolidated Standards of Reporting Trials (CONSORT; Moher et al., 2010) diagram for the study is shown in **Exhibit 1** and provides details of the flow of schools and teachers during the two years of the study. Characteristics of the 32 randomized schools are provided in **Exhibit 2**. Schools averaged an enrollment of 490 students, 27.6 full-time equivalent (FTE) teachers, a student-teacher ratio of 17.7, and a school population of 92.1% that is eligible for free lunch. There were no statistically significant differences between MFA and BAU schools in enrollment, teacher FTEs, numbers of students eligible for free lunch, or in proportion of students eligible for free lunch.

Teachers. An average of 50% of grade 4 and grade 5 teachers at each school agreed to participate (**Exhibit 3**). There were slightly more grade 4 teachers than grade 5 teachers, and most participating teachers were white and female, with some graduate-level education. The majority of teachers had at least six years of experience working with students with disabilities (SWD) but did not have formal SWD certification or a math education background. The average years of teaching was 11.56 in the MFA group and 13.47 in the BAU group.

As with any school-level trial, there was some mobility at the sub-cluster level. The team determined that any teacher who was in a study school at time of randomization (October 2015) but did not consent to participate or left the school is counted as part of sub-cluster attrition. Five teachers joined the sample of schools during the 2015–2016 year (see Exhibit 1). We consider these teachers to represent a low risk of introducing bias in the impact analyses. Our reasoning here is that teachers who were hired after randomization were unlikely to have had any awareness of whether the school was even participating in a study, much less the study condition. Teachers who joined the study schools in the 2016–2017 year are considered to be "late joiners," and any defense for including them in ITT estimates is less tenable.

We ran simple descriptive analyses and saw no meaningful differences between what we refer to as the "early joiner" teachers and the randomized group teachers. Hence, these teachers are included in the main impact analyses. Sensitivity analyses showed no meaningful difference in findings when the early joiner teachers are excluded.

Students. Parental consent statistics are shown in **Exhibit 4**. In Year 1, 2,242 students appeared on teacher rosters and 67% returned a parental consent form. Of the 1,481 who returned a consent form, at least 82% of parents provided consent for the various data collection aspects of the study. In Year 2, 1,735 students appeared on rosters and 59% returned a parental consent form. At least 85% of parents provided consent for the various data collection aspects of the study.

Demographic characteristics of students whose parents provided consent are shown in **Exhibit 5**. These data were provided by CPS and based on spring 2015 district rosters; of the 1,237 students with parental consent to release demographic data, 1,096 were matched to district administrative records. The majority of students were African American or Hispanic and economically disadvantaged. SWD were 11.3% of the sample, and English Language Learners (ELLs) made up 22.7% of the sample. The proportion of girls was slightly higher than that of boys (42.8% and 41.2%, respectively).

Challenges in Obtaining High Participation Rates. As the discussion above indicates, there were challenges in maintaining schools, recruiting teachers within schools, and engaging parents to allow their children to be part of the research. The study took place during a very challenging time in Chicago; the district was facing a big budget deficit, which resulted in school closures and severe cutbacks to expenditures. The teachers' union was at odds with the district, and the district was at odds with the state legislature. Morale was very low, as researchers at the Chicago Consortium on School Research recently reported (Gordon et al., 2018).

Anything beyond the usual was not welcomed. We found that in several schools, teachers did not make collection of student consent forms a priority and principals did not want to spend social capital they held with staff on study activities, and instead wanted to focus on maintaining the best possible climate during the labor dispute. We also believe that the low parental consent rates may have also been partially caused by the district strife, which received wide and daily coverage by the media.

Another challenge was that we had to conduct the PD on Saturdays because district administrators did not want teachers to be pulled from their classrooms during school hours. Although teachers were paid for their time, participation was voluntary, which undermined the gradewide participation of teachers and the collaboration between general and special education teachers. Within-school participation varied from one to two teachers at the school to most of the teachers in grades 4 and 5. Low morale also made teachers reluctant to agree to be observed and/or video-recorded.

Teacher transfers, retirements, layoffs, and firings also posed a challenge in maintaining our teacher sample between Years 1 and 2 of the study. We lost 12/46 (26%) MFA and 16/52 (31%) BAU teachers from spring 2016 to fall 2016. There were 19 teachers (*N* = 10 BAU; *N* = 9 MFA) who transferred from study schools and who were willing to provide us with data in Year 2. For this intent-to-treat (ITT) teacher sample, we collected teacher data, but not student data, in the second year of the study. There were 25 teachers who joined the study in fall 2016 (11 MFA; 14 BAU).[2] **Because RQ1, RQ2, and RQ3 focus on teacher outcomes at the end of the implementation year, we present results for Year 1. Year 2 results are considered exploratory and are not included in this brief.**

We had substantial year-to-year student turnover. Of the 2,242 students on a roster in Year 1, 592 were also on a roster in Year 2 (26%). Of the 592 that were on rosters in both years, 516 (87.2%) were fourth-graders in Year 1, and 76 (12.8%) were fifth-graders. Forty-three (7.3%) were fourth-graders again in Year 2, and 549 (92.7%) were fifth-graders in Year 2. Our impact sample of students who were in the study during grade 4 and grade 5 (Years 1 and 2) was further limited by parental consent status and ability to match the student to district administrative data. As a result, a fifth research question about the two-year effect of MFA on student achievement in mathematics could not be addressed because the sample of students who had parental consent and mathematics achievement data for the full span of the study was just over 100 (*N* = 43 BAU; *N* = 60 MFA).

We did examine cluster-level data to try to assess the two-year impact on student math achievement, but we cannot draw valid conclusions from the cluster-level analyses because:

- Any effect (Year 1 or Year 2) is diluted because not all the grade 4 and grade 5 teachers in an MFA school participated in the PD.
- The two-year MFA effect would be further diluted because 26% of the MFA teachers who participated in the MFA PD in Year 1 left the schools after spring 2016.
- It is unknown how many of the students who were in grade 5 in 2016–2017 were also at the schools in 2015–2016, when they were in grade 4.

Given the student sample loss between Years 1 and 2, the results presented here for RQ4 are limited to Year 1 findings. **Although we had attrition during the first year, the randomized school (i.e., cluster) sample is intact and internal validity of the study is preserved during Year 1. The Year 1 student-level impact analyses were done at the cluster level.**

Measures

Teacher Mathematical Content Knowledge. Twelve items from the Learning Mathematics for Teaching's item banks were used to construct a brief measure of teachers' content and pedagogical content knowledge in mathematics. The items covered number concepts and operations, geometry, and patterns functions and algebra (**Exhibit 6**). A total of 14 items comprised the mathematical knowledge for teaching (MKT) scale because Item #6 had multiple parts (6a, 6b, 6c). Cronbach alphas for the MKT scale ranged from 0.67 to 0.73 (**Exhibit 8**).

Teacher Comfort and Preparedness to Teach SWD. The teacher survey included two 11-item scales to measure teachers' self-reported comfort and preparedness to teach mathematics to diverse students,

[2] As shown in the CONSORT diagram, five "early joiners" (one MFA; four BAU) were added to the staff at five different schools in December 2015. Given the challenges CPS was experiencing with staffing, we believe it is implausible that these teachers joining their schools is related to MFA school status.

including those with disabilities. Item wordings are presented in **Exhibit 7**. Cronbach alphas ranged from 0.89 to 0.95 and are shown in **Exhibit 8**.

Classroom Assessment Scoring System (CLASS) Observation Rubric. The CLASS measures the quality of teacher-student interactions within four domains: emotional support, classroom organization, instructional support, and student engagement (Pianta, Hamre, & Mintz, 2012). Each of the domains is divided into dimensions of classroom quality. Observers typically watch a lesson for 15 minutes, taking notes on the specific behaviors they observe related to each of the CLASS dimensions. Scoring is completed at the dimension level using a 7-point scale, with the low range being a score of 1–2, the middle range 3–5, and the high range 6–7. The CLASS manual provides detailed information to help observers determine the specific score. The observer then watches the next 15 minutes and scores each of the dimensions again, repeating this cycle of observation and scoring until the end of the lesson. Lesson scores are created by averaging scores across all 15-minute cycles, and scores for teachers are averaged across lessons. Observations can be scored live or using video. **Exhibit 9** summarizes CLASS domains, dimensions, and indicators.

Student Mathematics Achievement.[3] The NWEA Measures of Academic Progress (MAP) assessment was the main measure of student achievement in mathematics. Information about the NWEA MAP's validity and reliability is presented in **Exhibit 10**.

Procedures

CLASS Training and Certification. In July 2015, 15 members of the research team participated in a two-day training on the Upper Elementary Classroom Assessment Scoring System (UE CLASS). Day 1 of the training involved reviewing each domain in depth and discussing the observable indicators of the dimensions that comprise each domain. Next, the participants viewed a video clip and live coded according to a specific UE CLASS domain. Participants discussed their assigned score for the teacher in the clip. Each video clip had a "master score." The trainer provided the master score as well as the rationale for the score. Day 2 included a review of the dimensions of each domain and live coding of videos using the entire instrument. Participants discussed their assigned scores for each training video and the trainer provided the master scores and rationales. The process continued over the course of the day to calibrate observation scores to the master coder. Within two weeks of the training, each research team member completed the online certification test. The online system included additional training videos to practice coding prior to taking the test. Criteria for passing the test were (1) coding within 1 point of master codes on 80% of the codes overall, and (2) demonstrating proficiency in each dimension by coding within 1 point of master codes on two out of five videos for each dimension. After initial certification, coders were tested and recertified annually via the CLASS online system.

Data Collection. The MKT, preparedness, and comfort scales were part of the teacher survey that was administered in fall 2015 and spring 2016. Because of limited resources, the study plan called for classroom observations of a random subsample of teachers (one grade 4 and one grade 5 teacher in each school). Classroom observations were intended to be videotaped, but in-person live observations were offered as an alternative to teachers who did not wish to be on camera. Forty teachers were observed in fall

[3] We also initially planned on using Partnership for Assessment of Readiness for College and Careers scores to understand student outcomes, but this is no longer policy-relevant since Illinois has discontinued use of this measurement system.

2015 and in spring 2016, but only 30 teachers were observed at both time points (i.e., had pretest and posttest data). End-of-year NWEA MAP assessments are administered every spring, so baseline data for our sample were from spring 2015 and posttest data were from spring 2016. Achievement and demographic data for students with parental consent were obtained from CPS in December 2016.

Computing MKT, Preparedness, and Comfort Scores (Rasch scaling). The teacher knowledge, preparedness, and comfort latent constructs were scaled under the Rasch unidimensional measurement model (Rasch, 1980) to obtain equal-interval teacher scale scores for use in standard statistical analyses (Bond & Fox, 2001; Waugh & Chapman, 2005; Wright, 1996). Each MKT item was coded as correct/incorrect (i.e., 1/0) and scored using Winsteps (Linacre, 2017). Pretest MKT items left unanswered were scored as incorrect. The Comfort and Preparedness responses are on 5-point Likert scales ranging from Not at all Comfortable to Very Comfortable and Not at all Prepared to Very Well Prepared, respectively; responses were also scaled using Winsteps (Linacre, 2017) under the Rasch Rating Scale model (Andrich, 1978a; 1978b; 1978c; Bond & Fox, 2001). For each scale and time point, assessment of model fit and unidimensionality was examined using measures of infit/outfit and Rasch factor analysis of residuals (Bond & Fox, 2001).

Computing CLASS Scores. Observers typically watch a lesson for 15 minutes, taking notes on the specific behaviors they observe related to each of the CLASS dimensions. Scoring is completed at the dimension level using a 7-point scale, with the low range being a score of 1–2, the middle range 3–5, and the high range 6–7. The CLASS manual provides detailed information to help observers determine the specific score. The observer then watches the next 15 minutes and scores each of the dimensions again, repeating this cycle of observation and scoring until the end of the lesson. Lesson scores are created by summing and averaging scores across dimensions for all 15-minute cycles.

Imputation. Multiple imputation procedures took into account the multilevel structure (i.e., teachers and students nested within schools) of the data (Enders, Mistler, & Keller, 2016; Keller & Enders, 2017) to generate 20 imputed data sets per outcome (Graham, Olchowski, & Gilreath, 2007). BLIMP software (Keller & Enders, 2017) was used to conduct Fully Conditional Specification (FCS; Enders, Keller, & Levy, 2017) and substantive model-compatible (SMC-FCS; Bartlett, Seaman, White, & Carpenter, 2014) multiple imputation using all analytically relevant variables and interactions, as well as a number of auxiliary variables related to each of the outcome variables. Model convergence diagnostics, including potential scale reduction (PSR; Gelman & Rubin, 1992) were reviewed to assess model adequacy. Multiple imputation was done for the three teacher outcomes (RQ1 and RQ2: MKT, preparedness, and comfort) and student achievement in mathematics (RQ4). Imputation was not done for RQ3 because classroom observations were done on a subsample of teachers. Because not all teachers were willing to be observed, the observations were not a true random subsample.

Results and Discussion

A summary of the impact analyses results is shown in **Exhibit 11**. We now turn to a review and discussion of the findings for each of the four research questions. Effect sizes for the teacher outcomes range from 0.43 to 0.98. Student achievement in mathematics was examined at both the cluster (school) level and individual student level. Effect sizes for the student outcomes ranged from 0.11 to 0.32. We also conducted exploratory analyses to test for the moderating effect of grade level.

Research Question #1. *Does participation in MFA PD, compared to business-as-usual (BAU) experiences of a control group, result in greater teacher **mathematical content knowledge** after the completion of the PD?*

The unimputed and imputed results for RQ1 are shown in **Exhibit 12**. There was a large baseline difference in MKT scores that favored the treatment group (*g* = 0.74 unimputed, 0.71 imputed). This may be an example of unhappy randomization or a function of the scoring procedure, where missing pretest MKT items were scored as zeros. There were more missing pretest MKT data in the control group, which might have been caused by lower motivation and/or lower effort by those teachers. Adjusted posttest means show that the MFA group mean is largely unchanged from fall to spring. The BAU group shows an increase from pretest to posttest, which could be attributed to a regression to the mean effect, or to control group teachers taking the posttest more seriously than they did the pretest. The overall effect favors the MFA group (effect sizes were 0.43 unimputed and 0.47 imputed), but we cannot conclude from these data that MFA improves teacher mathematical content knowledge.

Research Question #2. *Does participation in MFA PD, compared to BAU experiences of a control group, improve teachers' **comfort and preparedness** to teach mathematics to diverse students (including those with disabilities) after the completion of the PD?*

The unimputed and imputed results for RQ2 are shown in **Exhibit 13** (preparedness) and **Exhibit 14** (comfort). The pattern of results was the same for both scales: the MFA group reported lower levels of preparedness and comfort at the pretest (*g*s for preparedness were –0.36 unimputed, –0.35 imputed; and for comfort, *g*s were –0.25 unimputed and –0.25 imputed) and showed a steep increase from fall to spring. The opposite pattern was observed in the control group: the BAU group reported higher levels of preparedness and comfort at the pretest then showed a sharp decrease from fall to spring. Results were statistically significant. Effect sizes for preparedness were 0.54 (unimputed) and 0.58 (imputed). Effect sizes for comfort were 0.67 (unimputed) and 0.71 (imputed). Teachers who took the MFA PD appeared to have greatly increased their senses of preparedness and comfort in teaching SWD.

Research Question #3. *Does participation in MFA PD, compared to the BAU experiences of a control group, result in improved **mathematics classroom practice** after the completion of the PD?*

As noted above, only a small subsample of teachers agreed to be observed so we used a quasi-experimental design approach to examine this specific contrast. Although findings favor the MFA group, they do not meet What Works Clearinghouse (WWC) standards for establishing baseline equivalence. The impact table and plot shown in **Exhibit 15** indicate that MFA teachers scored higher in the **Emotional Support** domain than BAU teachers (*g* = 0.31). After controlling for pretest scores and teacher-level covariates, the effect size for the difference between groups was 0.98. Correlations between the four CLASS domains range from 0.68 to 0.83 at the pretest and 0.74 to 0.88 at the posttest. The Bonferroni correction for multiple comparisons renders the Emotional Support domain result not statistically significant at $p = 0.01$. Examination of group means on the dimensions that comprise the Emotional Support domain show that at the posttest, the MFA group was consistently higher in Positive Climate, Teacher Sensitivity, and Regard for Student Perspectives. This finding is consistent with the MFA PD's emphasis on building teachers' understanding of students' strengths and weaknesses.

Treatment group teachers also showed higher pretest means in Instructional Support (g = 0.41), Classroom Organization (g = 0.45), and Student Engagement (g = 0.66). Adjusted mean differences in posttest **Instructional Support** (**Exhibit 16**), **Classroom Organization** (**Exhibit 17**) and **Student Engagement** (**Exhibit 18**) were not statistically significant, but MFA teachers displayed higher adjusted posttest means than their BAU counterparts in the domain scores and in the descriptive posttest means for the dimension scores. Effect sizes were 0.69, 0.78, and 0.54 for Instructional Support, Classroom Organization, and Student Engagement, respectively. Recall that RQ3 involves smaller sample sizes because not all teachers were observed.

Research Question #4. *Does the use of an MFA approach in the classroom result in improved* ***student achievement in mathematics after one year of intervention exposure****?*

We examined MFA's impact on student achievement in mathematics in three ways: (1) a school-level analysis; (2) a student-level analysis; and (3) a student-level analysis that included grade level as a moderator.

Because parental consent rates were lower than desired, we conducted a **school (or cluster-level) aggregated analysis** to assess MFA's impact on student achievement on **all** grade 4 and grade 5 students at the 32 study schools. As shown in **Exhibit 19**, the adjusted posttest mean for the MFA schools was higher than that of the BAU schools. The effect size was 0.33, but not statistically significant, likely because of the small sample size. Nevertheless, we are encouraged by this finding, because the results favor the MFA group, even in a cluster analysis that dilutes the treatment effect (because the analysis includes students of nonparticipating grade 4 and grade 5 teachers at the treatment schools).

The **student-level analyses** (**Exhibit 20**) mirror the pattern shown in the cluster-level analysis. Although the results favor the treatment group (Hedges's g = 0.11 unimputed and 0.14 imputed), the differences were not statistically significant.

When **grade level was examined as a moderator**, we found different patterns between the grade 4 and the grade 5 samples. In **grade 4**, students whose teachers participated in the MFA PD had higher posttest NWEA MAP scores than students whose teachers were in the BAU group. Hedges's g was 0.20 (unimputed) and 0.26 (imputed); the impact analysis based on imputed data yielded a statistically significant result (**Exhibit 21**). In **grade 5**, there were very small mean differences between the MFA and BAU groups; indeed, the control group had slightly higher posttest means than the treatment group (**Exhibit 22**). These differences were not statistically significant.

The large impacts on teacher dispositions and the grade-level interaction on student achievement that we observed in our data indicate that teacher mediators and contextual factors may merit greater attention in PD theories of change. The results suggest that the pathway from PD to teacher practice may not necessarily be a linear progression; perhaps a more dynamic model can capture the relationships more accurately (**Appendix B**).

References

Andrich, D. (1978a). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement*, *2*(4), 581–594.

Andrich, D. (1978b). Rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561–573.

Andrich, D. (1978c). Scaling attitude items constructs and scored in the Likert tradition. *Educational and Psychological Measurement*, *38*(3), 665–680.

Ball, D. (1996). Teacher learning and the mathematics reforms: What we think we know and what we need to learn. *Phi Delta Kappan*, *77*(7), 500–508.

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

Bartlett, J. W., Seaman, S. R., White, I. R., & Carpenter, J. R. (2015). Multiple imputation of covariates by fully conditional specification: accommodating the substantive model. *Statistical Methods in Medical Research*, *24*(4), 462–487.

Birman, B., LeFloch, K. C., Klekotka, A., Ludwig, M., Taylor, J., Walters, K., Wayne, A., & Yoon, K.S. (2007). *State and local implementation of the No Child Left Behind Act, volume II—Teacher quality under NCLB: Interim report*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.

Boller, K., Atkins-Bennett, S., Malone, E. M., Baxter, G. P., & West, J. (2010). *Compendium of student, teacher, and classroom measures used in NCEE evaluations of educational interventions (Vol. I). Measures selection approaches and compendium development methods* (NCEE 2010-4012). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Bond, T. & Fox C. (2001). *Applying the Rasch model: Fundamental measurement in the health sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Brodesky, A. R., Gross, F. E., McTigue, A. S., & Palmer, A. (2007). A model for collaboration: Study groups are an effective way to plan math instruction for students with special needs. *Educational Leadership, 64*(5). Alexandria, VA: ASCD. Available online at http://www.ascd.org/publications/educational-leadership/feb07/vol64/num05/A-Model-for-Collaboration.aspx

Brown, J., Bransford, A. L., & Cocking, R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academies Press.

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

Cohen, D. K., & Hill, H. C. (1998). *Instructional policy and classroom performance: The mathematics reform in California* (CPRE RR-39). Philadelphia, PA: 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. New Brunswick, NJ: Rutgers, State University of New Jersey.

Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work*. San Francisco, CA: Jossey-Bass.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective Teacher Professional Development*. Palo Alto, CA: Learning Policy Institute.

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.

Enders, C. K., Mistler, S. A., & Keller, B. T. (2016). Multilevel multiple imputation: A review and evaluation of joint modeling and chained equations imputation. *Psychological Methods*, *21*(2), 222–240.

Enders, C.K., Keller, B. T., & Levy, R. (2017). A fully conditional specification approach to multilevel imputation of categorical and continuous variables. *Psychological Methods*. Advance online publication. http://dx.doi.org/10.1037/met0000148

Espin, C., Shin, J., & Busch, T. (2000). *Formative evaluation. Current Practice Alert No. 3*. Reston, VA: Council for Exceptional Children. Retrieved from www.teachingld.org/pdf/Alert3.pdf

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, *53*(3), 199–208.

Fuchs, L. S., & Fuchs, D. (2002). *What is scientifically-based research on progress monitoring?* (Technical Report). Nashville, TN: Vanderbilt University.

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.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. S*tatistical Science*, 457–472.

Gordon, M. F., de la Torre, M., Cowhy, J. R., Moore, P. T., Sartain, L. S., & Knight, D. (2018). *School closings in Chicago: Staff and student experiences and academic outcomes*. Chicago, IL: University of Chicago Consortium on School Research.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, *8*, 206–213.

Hamre, B. K. (2018, January). Effective strategies for training and maintaining high quality use of measures in the field: Lessons from the Classroom Assessment Scoring System (CLASS). Presented at the meeting of the Institute of Educational Sciences Principal Investigators, Arlington, VA.

Keller, B. T., & Enders, C. K. (2017). *BLIMP Software Manual (Version 1.0)*. Los Angeles, CA.

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

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

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.

Linacre, J. M. (2017). *Winsteps® Rasch measurement computer program*. Beaverton, Oregon: Winsteps.com

Loucks-Horsley, S., Hewson, P. W., Love, N., & Stiles, K. E. (1998). *Designing professional development for teachers of science and mathematics*. Thousand Oaks, CA: Corwin Press.

Loucks-Horsley, S., Love, N., Stiles, K. E., Mundry, S., & Hewson, P. (2010). *Designing professional development for teachers of science and mathematics* (2nd ed.). Thousand Oaks, CA: Corwin Press.

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

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

Moher, D., Hopewell, S., Schulz, K.F., Montori, V., Gotzsche, P.C., Devereaux, P.J., Elbourne, D., Egger, M., & Altman, D.G. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ*, *340*, 1-28. doi: 10.1136/bmj.c869

Pianta, R. C., Hamre, B. K., & Mintz, S. L. (2012). *Classroom Assessment Scoring System—Secondary Manual*. Charlottesville, VA: Teachstone.

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

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests: Expanded edition with foreword and afterword*. Chicago, IL: The University of Chicago Press.

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.

Taylor, J. A., Roth, K., Wilson, C. D., Stuhlsatz, M. A. M., & Tipton, E. (2017) The effect of an analysis-of-practice, videocase-based, teacher professional development program on elementary students' science achievement. *Journal of Research on Educational Effectiveness*, *10*(2), 241–271, DOI: 10.1080/19345747.2016.1147628

Waugh, R., & Chapman, E. (2005). An analysis of dimensionality using Factor analysis (True-Score theory) and Rasch measurement: What is the difference? Which method is better? *Journal of Applied Measurement*, *6*(1), 80–99.

Wei, R. C., Darling-Hammond, L., Andree, A., Richardson, N., Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad*. Dallas, TX: National Staff Development Council. Retrieved from https://learningforward.org/docs/default-source/pdf/nsdcstudytechnicalreport2009.pdf

Wei, R. C., Darling-Hammond, L., & Adamson, F. (2010). *Professional development in the United States: Trends and challenges*. Dallas, TX: National Staff Development Council. Retrieved from http://npsct.org/pdf/resources/Topics_by_Category/Professional%20Development%20and%20Calendar%20Information/Details/Professional%20Learning%20in%20the%20US%20-%20Trends.pdf

Wright, B. (1996). Comparing Rasch measurement and Factor analysis. *Structural Equation*, *3*(1), 3–24.

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007-No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.

Exhibit 1. CONSORT diagram for MFA efficacy study

October 2015: Randomization			
	# of schools	# of teachers	# of participating teachers
BAU	16	98	49
MFA	16	93	47
Total	32	191	96
Fall 2015: If early joiners are treated as functionally "there" at random assignment			
	# of schools	# of teachers	# of participating teachers
BAU	16	102	53
MFA	16	94	48
Total	32	196	101
Late fall 2015, removing teachers from 2 BAU and 1 MFA attrited schools, and including joiners			
	# of schools	# of teachers	# of participating teachers
BAU	14	95	53 (includes 4 joiners)
MFA	15	91	48 (includes 1 joiner)
Total	29	186	101
Spring 2016			
	# of schools	# of teachers	# of participating teachers
BAU	14	95	52 (lost 1 joiner)
MFA	15	91	46 (lost 2 joiners)
Total	29	186	98
Fall 2016			
	# of schools	# of teachers	# of participating teachers
BAU	13	111	50 (lost 16; added 14 joiners)
MFA	13	96	49 (lost 12; added 11 joiners and 4 flippers)
Total	26	207	99
Spring 2017			
	# of schools	# of teachers	# of participating teachers
BAU	13	111	45
MFA	13	96	48
Total	26	207	93

Notes. Impact analyses are reported for Year 1 only and included early joiners in the analysis sample. Numbers for Year 2 are provided to document the attrition between Years 1 and 2. A “flipper” is a teacher who was present at randomization, decided not to participate during Year 1, but chose to participate in Year 2.

Exhibit 2. Characteristics of MFA schools (randomized sample)

School	Magnet	Title I School	Title 1 Schoolwide	Enrollment	FTE Teachers	Student-Teacher Ratio	Free Lunch %
BAU-01	No	Yes	Yes	452	26.19	17.3	96.2%
BAU-02	No	Yes	Yes	478	28.7	16.7	98.5%
BAU-03	No	Yes	Yes	478	28.4	16.8	92.5%
BAU-04	No	Yes	Yes	360	18.73	19.2	91.7%
BAU-05	No	Yes	Yes	299	18.7	16	95.7%
BAU-06	No	Yes	Yes	328	20.93	15.7	95.4%
BAU-07	No	Yes	Yes	329	19.49	16.9	97.9%
BAU-08	Yes	Yes	No	645	37.26	17.3	50.7%
BAU-09	No	Yes	Yes	817	40.98	19.9	95.8%
BAU-10	No	Yes	Yes	440	20.59	21.4	97.5%
BAU-11	No	Yes	Yes	953	53.01	18	97.2%
BAU-12	No	Yes	Yes	678	36.41	18.6	98.4%
BAU-13	No	Yes	Yes	152	10.06	15.1	96.1%
BAU-14	No	No	N/A	561	33.33	16.8	44.9%
BAU-15	No	Yes	Yes	280	14.1	19.9	93.9%
BAU-16	No	Yes	Yes	543	35.65	15.2	100.0%
MFA-01	No	Yes	Yes	535	31.5	17	96.6%
MFA-02	No	Yes	Yes	377	21.22	17.8	96.3%
MFA-03	No	Yes	Yes	473	23.1	20.5	95.3%
MFA-04	No	Yes	Yes	991	54.08	18.3	86.5%
MFA-05	No	Yes	Yes	295	19.72	15	95.9%
MFA-06	No	Yes	Yes	234	15.44	15.2	100.0%
MFA-07	No	Yes	Yes	367	21.94	16.7	82.3%
MFA-08	No	Yes	Yes	644	38.94	16.5	89.9%
MFA-09	No	Yes	Yes	856	43.27	19.8	99.3%
MFA-10	No	Yes	Yes	215	13.71	15.7	100.0%
MFA-11	No	Yes	Yes	364	17.76	20.5	90.7%
MFA-12	Yes	Yes	Yes	506	26.04	19.4	98.4%
MFA-13	No	Yes	Yes	536	31.37	17.1	82.5%
MFA-14	No	Yes	Yes	539	29.49	18.3	97.8%
MFA-15	No	Yes	Yes	567	32.52	17.4	98.2%
MFA-16	No	Yes	Yes	387	20.1	19.3	95.1%

Source: National Center for Education Statistics, 2015–2016 Common Core of Data.

Exhibit 3. Characteristics of teachers at MFA schools (from fall 2015 pretest data)

	MFA		BAU		Total	
	n	%	*n*	%	*n*	%
Total number of teachers	**45**	**46.4%**	**52**	**53.6%**	**97**	**100.0%**
Grade 4	23	51.1%	25	48.1%	48	49.5%
Grade 5	19	42.2%	23	44.2%	42	43.3%
Grade 4 & 5	3	6.7%	4	7.7%	7	7.2%
Minority Race	17	37.8%	12	23.1%	29	29.9%
Female	34	75.6%	32	61.5%	66	68.0%
Male	8	17.8%	4	7.7%	12	12.4%
Gender Not Reported	2	4.4%	1	1.9%	3	3.1%
SWD Teacher	14	31.1%	11	21.2%	25	25.8%
SWD Certification	10	22.2%	5	9.6%	15	15.5%
At least 6 years of experience w/SWD	29	64.4%	21	40.4%	50	51.5%
Math Education Background	8	17.8%	5	9.6%	13	13.4%
Some Graduate Education	34	75.6%	34	65.4%	68	70.1%
Mean Years Teaching	11.56		13.47			

Exhibit 4. Parental consent statistics

	Year 1		Year 2	
	n	%	*n*	%
Students on Roster	2,242	100%	1,735	100%
Returned Consent Form	1,481	67%	1,028	59%
Missing Consent Form	744	33%	707	41%
Of those returned				
Survey Consent	1,349	91%	951	93%
Video Consent	1,213	82%	873	85%
Assessment Consent	1,256	85%	887	86%
Demographic Consent	1,237	84%	886	86%

Exhibit 5. Characteristics of students at MFA schools in fall 2015 whose parents consented to providing demographic data (from CPS spring 2015 data)

	MFA		BAU		Total	
	n	%	n	%	n	%
Total number of students	**536**	**48.9%**	**560**	**51.1%**	**1,096**	**100.0%**
Female	240	44.8%	229	40.9%	469	42.8%
Male	207	38.6%	245	43.8%	452	41.2%
Asian	11	2.1%	2	0.4%	13	1.2%
African American	173	32.3%	229	40.9%	402	36.7%
Hispanic	234	43.7%	228	40.7%	462	42.2%
Multi	3	0.6%	1	0.2%	4	0.4%
Native American	1	0.2%	0	0.0%	1	0.1%
White	25	4.7%	14	2.5%	39	3.6%
SWD	58	10.8%	66	11.8%	124	11.3%
ELL	141	26.3%	108	19.3%	249	22.7%
Economically Disadvantaged	429	80.0%	434	77.5%	863	78.7%

Note. Because parental consent was required to obtain demographic data from the district, we are unable to describe how similar/different the demographics are between the students with parental consent and those without parental consent.

Exhibit 6. Mathematical Knowledge for Teaching (MKT) Scale

Please do not spend more than 1-2 minutes on any question in this section. Imagine you are responding to real classroom situations and select the answer that most closely matches what you would do, say, or answer at that moment.

1. (EL_GEO-CK_2004A_form)

Mr. Nager writes the following statement on the board:

The length and width of a rectangular swimming pool are each doubled, while the depth remains the same.

He asks his students to make mathematical statements about this pool. Which of the following student claims is true? (Mark ONE answer.)

a) It takes twice as much paint to paint the bottom.
b) It takes twice as much paint to paint the four walls.
c) It takes twice as much water to fill the pool.
d) All of the above.
e) None of the above.
f) I'm not sure.

2. (EL_NCOP-CK_2001A_form)
Ms. Harris was working with her class on divisibility rules. She told her class that a number is divisible by 4 if and only if the last two digits of the number are divisible by 4. One of her students asked her why the rule for 4 worked. She asked the other students if they could come up with a reason, and several possible reasons were proposed. Which of the following statements comes closest to explaining the reason for the divisibility rule for 4? (Mark ONE answer.)

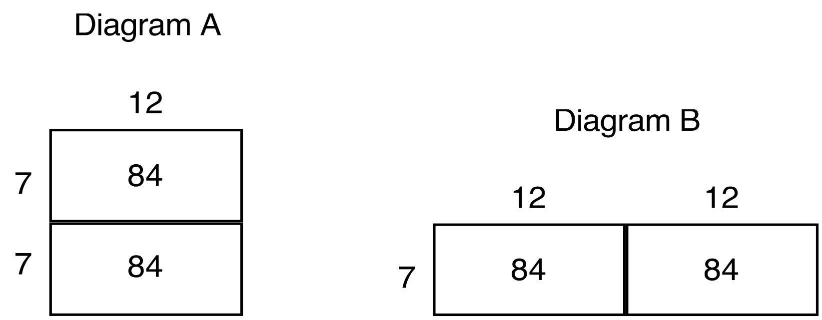
a) Four is an even number, and odd numbers are not divisible by even numbers.
b) The number 100 is divisible by 4 (and also 1000, 10,000, etc.).
c) Every other even number is divisible by 4, for example, 24 and 28 but not 26.
d) It only works when the sum of the last two digits is an even number.

3. (EL_NCOP-CK_2004A_form)

Luanne suggested the following method for multiplying 14 by 12:

> I know that 7 times 12 is 84, so to get 14 times 12, I double 84, which is 168.

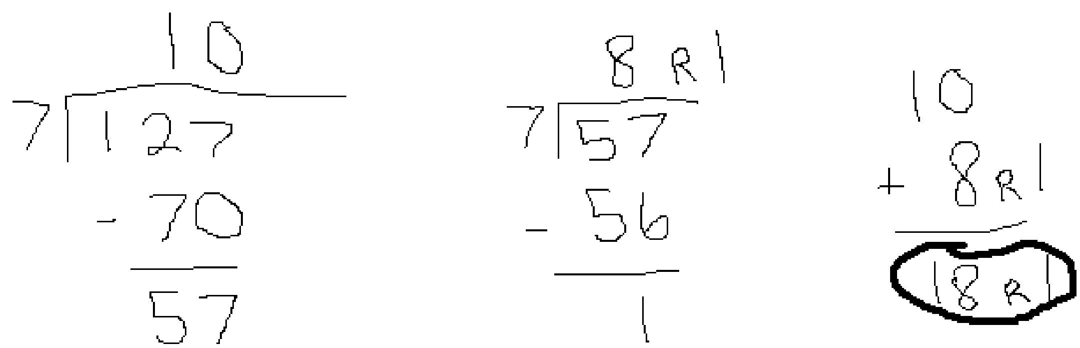
Of the following diagrams, which BEST illustrates Luanne's method? (Mark ONE answer.)



a) Diagram A only
b) Diagram B only
c) Both diagrams represent Luanne's method equally well.
d) Neither diagram represents Luanne's method well.
e) I'm not sure.

4. (EL_NCOP-CK_2004B_form)
Ms. Barber was reviewing her students' division homework and saw that Chad used the following non-standard approach to divide 127 by 7:

127 divided by 7.



What is true about Chad's approach?

a) His approach is not mathematically valid; it is a coincidence that his answer is correct.
b) His approach is not mathematically valid because he subtracted 70 from 127 instead of subtracting 7 from 12.
c) His approach is mathematically valid, but could be inefficient with large dividends.
d) His approach is mathematically valid, but only works with single-digit divisors.
e) I'm not sure.

5. (EL_NCOP-KCS_2001A_form)

You are working individually with Bonny, and you ask her to count out 23 checkers, which she does successfully. You then ask her to show you how many checkers are represented by the 3 in 23, and she counts out 3 checkers. Then you ask her to show you how many checkers are represented by the 2 in 23, and she counts out 2 checkers. What problem is Bonny having here? (Mark ONE answer.)

a) Bonny doesn't know how large 23 is.
b) Bonny thinks that 2 and 20 are the same.
c) Bonny doesn't understand the meaning of the places in the numeral 23.
d) All of the above.

6. When learning about multi-digit subtraction, Mrs. Bisson's class encounters the following problem:

$$\begin{array}{r} {\scriptstyle 2\ 9\ 16} \\ \cancel{3}\cancel{0}\cancel{6} \\ \underline{-\ 97} \\ 209 \end{array}$$

A student answers "209" by completing the problem on the board as shown above. Mrs. Bisson wants to ensure that students understand <u>why the procedure works</u> and asks students to explain this answer. Which explanation(s) should she feel comfortable accepting as evidence a student understands why the procedure works? (Circle YES, NO or I'M NOT SURE for each.)

	Yes	No	I'm not sure
a) "You can't take 7 from 6, so you cross out the 0 and make it a 9, and the 6 becomes a 16, and then cross out the 3 and it becomes a 2. Then you take away. 16 take away 7 is 9, 9 take away 9 is 0, and you just have 2."	1	2	3
b) "She regrouped 306 to be 2 hundreds, 9 tens, and 16 ones. That's the same amount as 306. Then she could do the problem. She took away 7 from 16, and she took away 9 from 9."	1	2	3
c) "She borrowed from the tens place to make the 6 a 16. But since it was a zero, she had to borrow again from the hundreds place, making the three a two. Then she just subtracted."	1	2	3

7. (EL_NCOP-KCS_2001C_form)

Sometimes it is difficult to figure out what students are talking about when they are working on problems. One of Mrs. Padamsee's students was overheard saying the following: "Seven times five is thirty-five, forty-two, forty-nine, fifty-six." What problem was the student <u>most likely</u> trying to solve? (Mark ONE answer.)

a) 35 + 21
b) 7 x 8
c) 28 + 28
d) 56 ˎ 7
e) I'm not sure.

8. (EL_NCOP-KCS_2001C_form)

Mrs. Jackson is getting ready for the state assessment and is planning mini-lessons for students focused on particular difficulties that they are having with multiplication computation. To target her instruction more effectively, she wants to work with groups of students who are making the same kind of error, so she looks at a recent quiz to see what they tend to do. She sees the following three student mistakes:

$$\text{I)}\quad \begin{array}{r} 24 \\ \times 12 \\ \hline 48 \\ 24 \\ \hline 72 \end{array} \qquad \text{II)}\quad \begin{array}{r} {\scriptstyle 2} \\ 503 \\ \times\ 27 \\ \hline 3501 \\ 10060 \\ \hline 13561 \end{array} \qquad \text{III)}\quad \begin{array}{r} {\scriptstyle 1} \\ 112 \\ \times\ 35 \\ \hline 560 \\ 336 \\ \hline 896 \end{array}$$

Which have the same kind of error? (Mark ONE answer.)

a) I and II
b) II and III
c) I and III
d) I, II, and III
e) I'm not sure.

9. Mr. Lewis was surprised when one of his students came up with a new procedure for subtraction (pictured below), and he wondered whether it would always work. He showed it to Ms. Braun, next door, and asked her what she thought.

$$\begin{array}{r} 37 \\ -19 \\ \hline -2 \\ 20 \\ \hline 18 \end{array}$$

What is true about this student's approach to the problem? (Circle ONE answer.)

a) The procedure works for this problem but would not work for all numbers.
b) This does not make sense mathematically.
c) This would work for all numbers.
d) This procedure only works in special cases.

10. (EL_PFA-CK_2001A_form)

It was Sally's birthday. Mr. Siegel and Sally made up a math problem for the class:

Sally is exactly twice as old as her brother. When will she be twice as old as him again?

The class generated the following ideas. Which of the following statements would you accept as correct? (Mark ONE answer.)

a) It will happen every two years.
b) It depends on Sally's age.
c) It will happen when she is twice as old as she is now.
d) It will never happen again.

11. (EL_PFA-CK_2001B_form)
Ms. Yolanta had her students cut rectangles out of paper to investigate area and perimeter. She posed the question: "If I start with a rectangle and make a new rectangle by doubling its length and halving its width, how does the area change?" Students volunteered many ideas. Which of their ideas about the area of the rectangles is true? (Mark ONE answer.)

a) The area always changes.
b) The area always stays the same.
c) The area sometimes changes.
d) It's not possible to tell from this information.

12. Mr. Hosko was wondering what it meant to say that division by 0 is *undefined*. He asked his colleague, Mrs. King, what she thought. Which of the following <u>best</u> explains this? (Circle ONE answer.)

a) Division by 0 is undefined because you cannot do it.
b) Division by 0 is undefined because you cannot make 0 groups of something.
c) Division by 0 is undefined in school curricula because college-level mathematics is needed to do this calculation.
d) Division by 0 is undefined because there is no single answer that when multiplied by the divisor 0 gives the original number.
e) Division by 0 is undefined because every number divided by 0 equals 0.

Exhibit 7. Preparedness and Comfort Scales

How well **prepared** do you feel with the following? (1 = not at all prepared; 7 = very well prepared)

How **comfortable** do you feel with the following? (1 = not at all comfortable; 7 = very comfortable)

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

Exhibit 8. Standardized Cronbach alphas for teacher survey scales

Scale	Fall 2015	Spring 2016	Fall 2016 (Joiners only)	Spring 2017
Mathematical Content Knowledge (MKT)	0.73 (*n* = 88)	0.67 (*n* = 94)	0.69 (*n* = 23)	0.69 (*n* = 90)
Preparedness to teach mathematics to students with disabilities	0.92 (*n* = 86)	0.94 (*n* = 93)	0.89 (*n* = 22)	0.92 (*n* = 89)
Comfort with teaching math to students with disabilities	0.93 (*n* = 83)	0.95 (*n* = 93)	0.92 (*n* = 22)	0.92 (*n* = 89)

Exhibit 9. CLASS domains, dimensions, and indicators (from Hamre, 2018)

Exhibit 10. Reliability and validity information for NWEA MAP Assessment

Internal consistency reliability coefficient alphas for mathematics ranged from 0.92 to 0.96. **Test-retest reliability** coefficients for mathematics ranged from 0.77 to 0.94 (Boller, Atkins-Bennett, Malone, Baxer, & West, 2010).	**Construct/concurrent validity:** Mathematics, reading, and language usage MAP scores correlated with the Iowa Test of Basic Skills, with coefficients ranging from 0.74 to 0.84. **Predictive validity:** The correlation coefficient for different versions of this assessment taken by the same students in spring and fall was 0.85 for mathematics (Malone et al., 2010).

Exhibit 11. Summary of impact analyses results

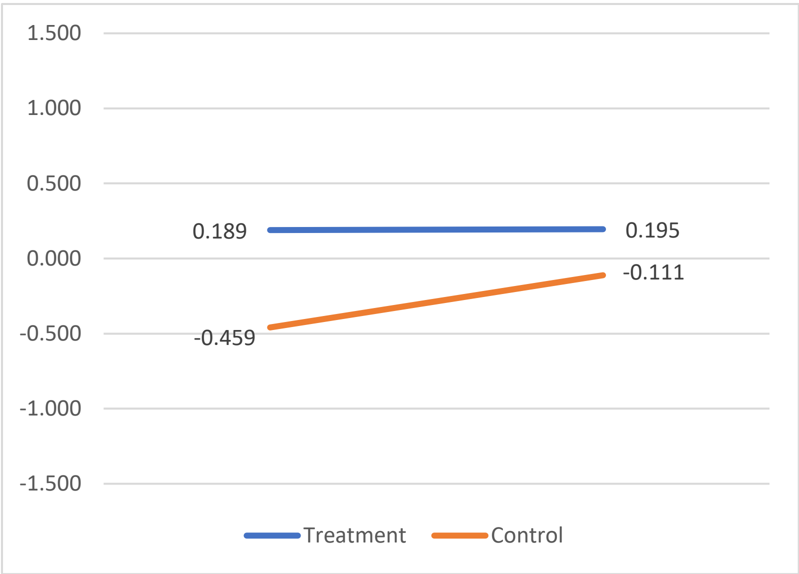
Outcome at the end of Year 1	Unimputed		Imputed	
	Hedges's *g*	***p*-value of treatment indicator**	**Hedges's *g***	***p*-value of treatment indicator**
Teacher Knowledge (MKT)	0.43	0.11	0.47	0.07
Teacher Preparedness	0.54	0.01	0.58	0.04
Teacher Comfort	0.67	0.01	0.71	0.01
CLASS Emotional Support	0.98	0.04	N/A	N/A
CLASS Instructional Support	0.69	0.08	N/A	N/A
CLASS Classroom Organization	0.76	0.06	N/A	N/A
CLASS Student Engagement	0.54	0.33	N/A	N/A
NWEA School Level	0.33	0.09	N/A	N/A
NWEA Student Level	0.11	0.39	0.14	0.19
NWEA Student Level – Grade 4	*0.20*	*0.16*	*0.20*	*0.03*
NWEA Student Level – Grade 5	*–0.06*	*0.65*	*–0.04*	*0.74*

Note. Positive values for Hedges's *g* indicate results that favor the MFA group.

Exhibit 12. Teacher Mathematical Content Knowledge (MKT) – Year 1 Impact Analyses

UNIMPUTED

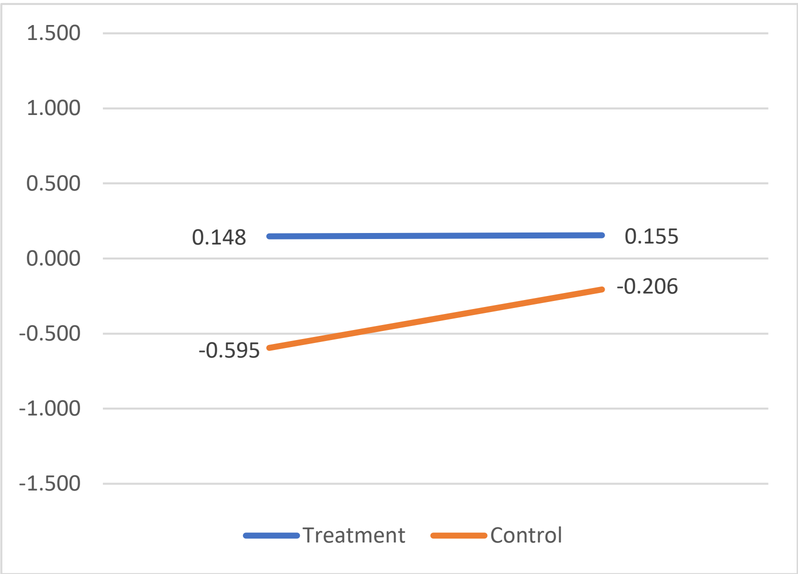
Group	Fall 2015 Pretest Mean Score	Fall 2015 Pretest SD	Spring 2016 Posttest Unadj. Mean Score	Spring 2016 Posttest Adj. Mean Score	Spring 2016 Posttest SD	Hedges' *g*	*p*-value
MFA	0.189	0.624	0.269	0.195	0.649	0.432	0.110
BAU	–0.459	1.059	–0.177	–0.111	0.757		



Note. Single-level Hierarchical Generalized Linear models/modeling (HGLH) with school dummies and baseline MKT (no other predictors were significant), using *n* = 42 BAU teachers and *n* = 43 MFA teachers. The baseline difference for the analytic sample used to estimate the adjusted mean scores below was computed as *g* = 0.74 (Unsatisfied).

IMPUTED

Group	Fall 2015 Pretest Mean Score	Fall 2015 Pretest SD	Spring 2016 Posttest Unadj. Mean Score	Spring 2016 Posttest Adj. Mean Score	Spring 2016 Posttest SD	Hedges's *g*	*p*-value
MFA	0.148	0.668	0.232	0.155	0.709	0.471	0.074
BAU	–0.595	1.286	–0.273	–0.206	0.812		

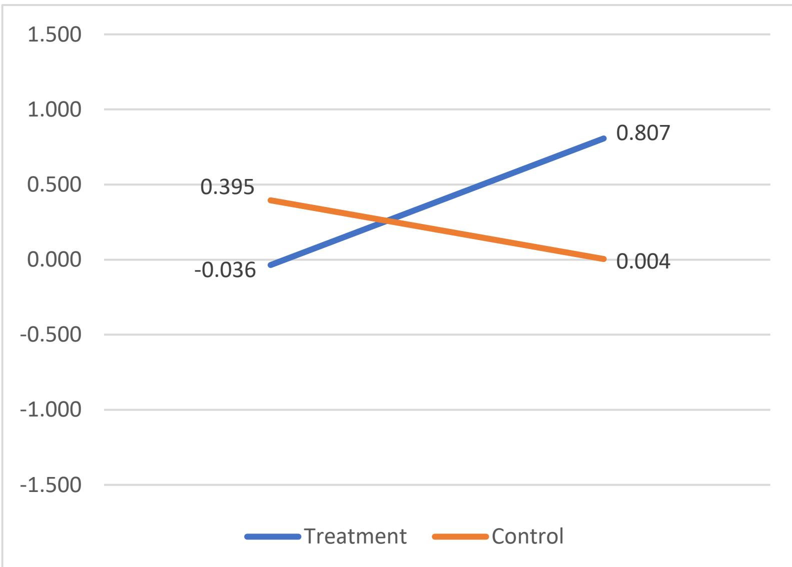


Note. Single-level HGLH with school dummies and baseline MKT (all other predictors were not significantly related to post MKT). The baseline difference for the analytic sample used to estimate the adjusted mean scores below was computed as *g* = 0.71 (Unsatisfied).

Exhibit 13. Teacher Preparedness in Teaching Students with Disabilities – Year 1 Impact Analyses

UNIMPUTED

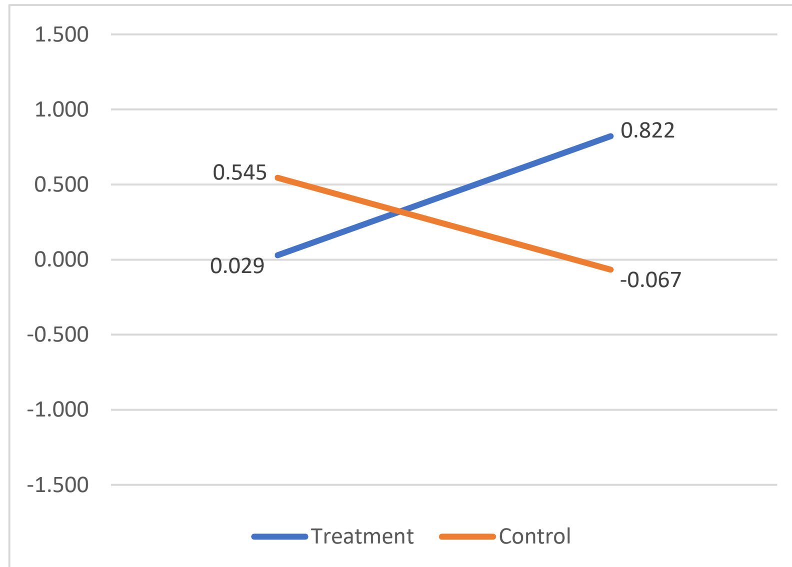
Group	Fall 2015 Pretest Mean Score	Fall 2015 Pretest SD	Spring 2016 Posttest Unadj. Mean Score	Spring 2016 Posttest Adj. Mean Score	Spring 2016 Posttest SD	Hedges's *g*	*p*-value
MFA	–0.036	1.193	0.748	0.807	1.322	0.541	0.013
BAU	0.395	1.180	0.060	0.004	1.623		



Note. Multilevel model with the Preparedness dependent variable, including the treatment indicator, the baseline Preparedness variable, the enjoy teaching math predictor, 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 MFA teachers. The baseline difference for the analytic sample used to estimate the adjusted mean scores below was computed as g = –0.36 (Unsatisfied).

IMPUTED

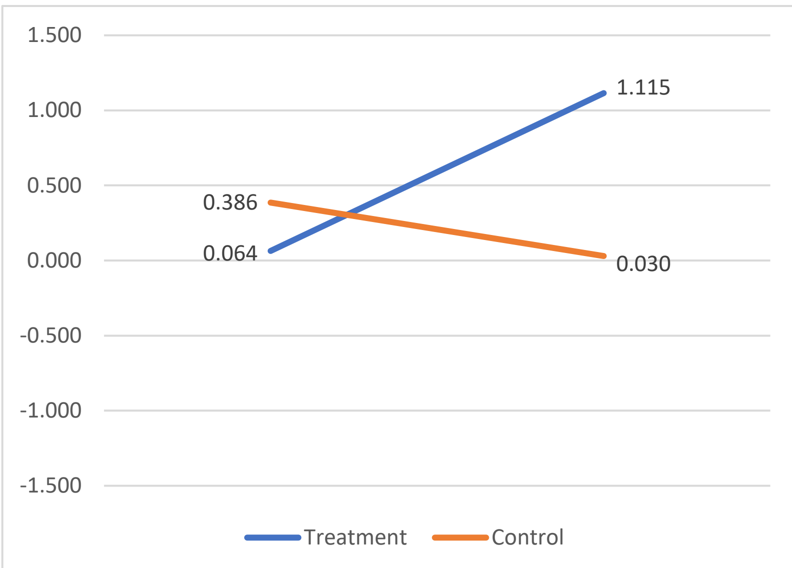
Group	Fall 2015 Pretest Mean Score	Fall 2015 Pretest SD	Spring 2016 Posttest Unadj. Mean Score	Spring 2016 Posttest Adj. Mean Score	Spring 2016 Posttest SD	Hedges's *g*	*p*-value
MFA	0.029	1.234	0.811	0.822	1.425	0.583	0.035
BAU	0.545	1.645	–0.087	–0.067	1.603		



Note. Multilevel model with baseline Preparedness and the enjoy teaching math predictor, and only a random intercept. The baseline difference for the analytic sample used to estimate the adjusted mean scores below was computed as g = –0.35 (Unsatisfied; which means the control group had a higher baseline mean).

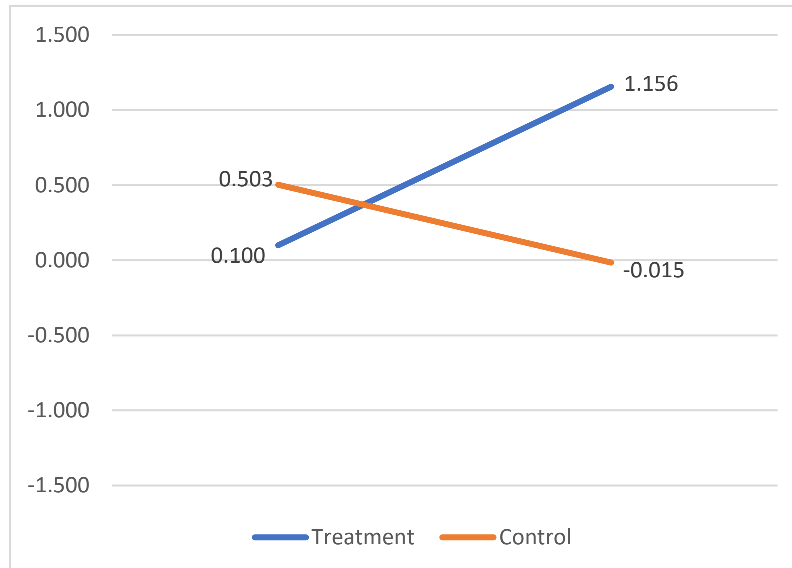
Exhibit 14. Teacher Comfort in Teaching Students with Disabilities – Year 1 Impact Analyses

UNIMPUTED							
	Fall 2015 Pretest		Spring 2016 Posttest				
Group	Mean Score	SD	Unadj. Mean Score	Adj. Mean Score	SD	Hedges's *g*	*p*-value
MFA	0.064	1.297	1.088	1.115	1.442	0.666	0.005
BAU	0.386	1.235	0.051	0.030	1.800		



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 MFA teachers. The baseline difference for the analytic sample used to estimate the adjusted mean scores below was computed as *g* = –0.25 (Unsatisfied).

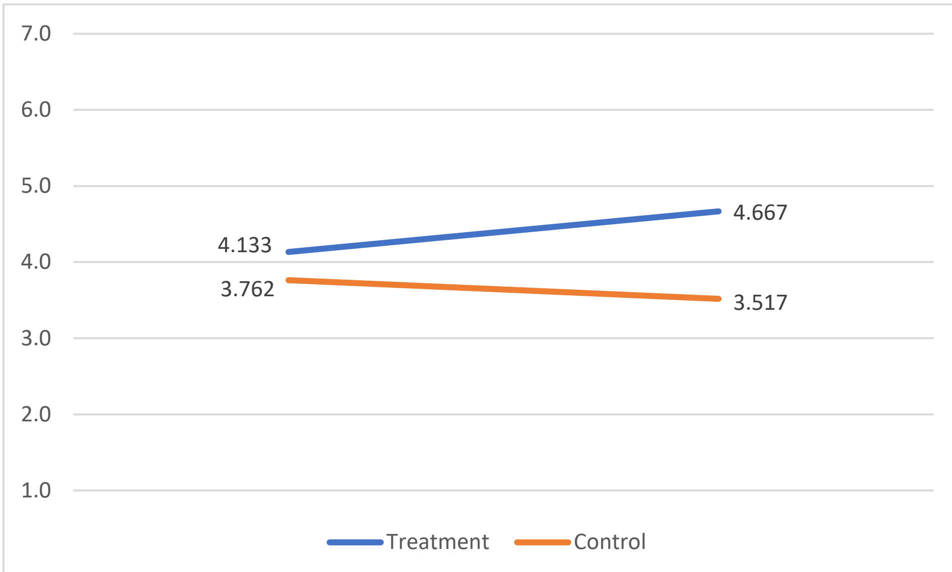
IMPUTED							
	Fall 2015 Pretest		Spring 2016 Posttest				
Group	Mean Score	SD	Unadj. Mean Score	Adj. Mean Score	SD	Hedges's *g*	*p*-value
MFA	0.100	1.305	1.143	1.156	1.525	0.712	0.009
BAU	0.503	1.795	–0.037	–0.015	1.738		



Note. Multilevel model with baseline Comfort and the enjoy teaching math predictor, and a random intercept only. The baseline difference for the analytic sample used to estimate the adjusted mean scores below was computed as *g* = –0.25 (Unsatisfied; which means the control group had a higher baseline mean).

Exhibit 15. CLASS Emotional Support Domain – Year 1 Impact Analyses

Group	Fall 2015 Pretest		Spring 2016 Posttest			Hedges's *g*	*p*-value
	Mean Score	SD	Unadjusted Mean Score	Adjusted Mean Score	SD		
MFA	4.133	1.066	4.519	4.667	1.080	0.982	0.037
BAU	3.762	1.297	3.821	3.517	1.181		

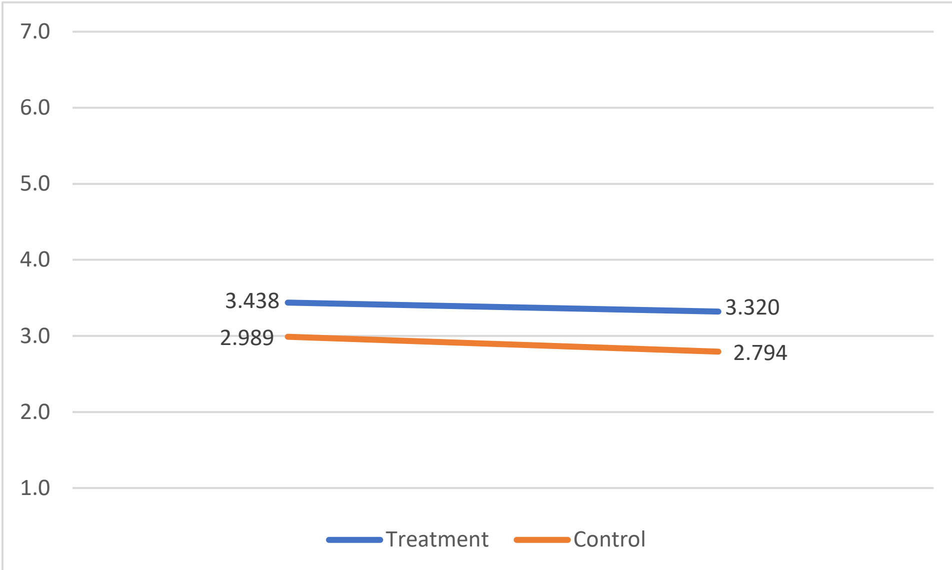


Note. The final adjusted multilevel model for Emotional Support included Baseline Emotional Support, SWD Experience (dichotomized), teacher racial minority indicator, Motivation for MFA PD, and Years Teaching Experience. The baseline difference for the analytic sample used to estimate the adjusted mean scores below was computed as $g = 0.31$ (Unsatisfied). Correlations between the four CLASS domains range from 0.68 to 0.83 at the pretest and 0.74 to 0.88 at the posttest. The Bonferroni correction for multiple comparisons renders the Emotional Support domain result not statistically significant at $p = 0.01$. Descriptive statistics for the Emotional Support dimensions are shown below (posttest means do not adjust for covariates).

CLASS Emotional Support Dimension	MFA Treatment				BAU Control			
	N	Pretest Mean (SD)	*N*	Posttest Mean (SD)	*N*	Pretest Mean (SD)	*N*	Posttest Mean (SD)
Positive Climate Relationships; positive affect; positive communications; respect	18	4.13 (1.09)	22	4.23 (1.20)	19	4.39 (1.39)	17	3.58 (1.13)
Teacher Sensitivity Awareness; responsiveness to academic and social/emotional needs and cues; effectiveness in addressing problems; student comfort	18	4.69 (1.33)	22	5.04 (1.13)	19	4.76 (1.58)	17	4.42 (1.35)
Regard for Student Perspectives Flexibility and student focus; connections to real life; support for autonomy and leadership; meaningful peer interactions	18	2.97 (1.11)	22	3.66 (1.18)	19	3.14 (1.21)	17	3.10 (0.95)

Exhibit 16. CLASS Instructional Support Domain – Year 1 Impact Analyses

Group	Fall 2015 Pretest		Spring 2016 Posttest			Hedges's *g*	*p*-value
	Mean Score	SD	Unadjusted Mean Score	Adjusted Mean Score	SD		
MFA	3.438	1.160	3.482	3.320	0.809	0.690	0.084
BAU	2.989	0.944	2.667	2.794	0.643		

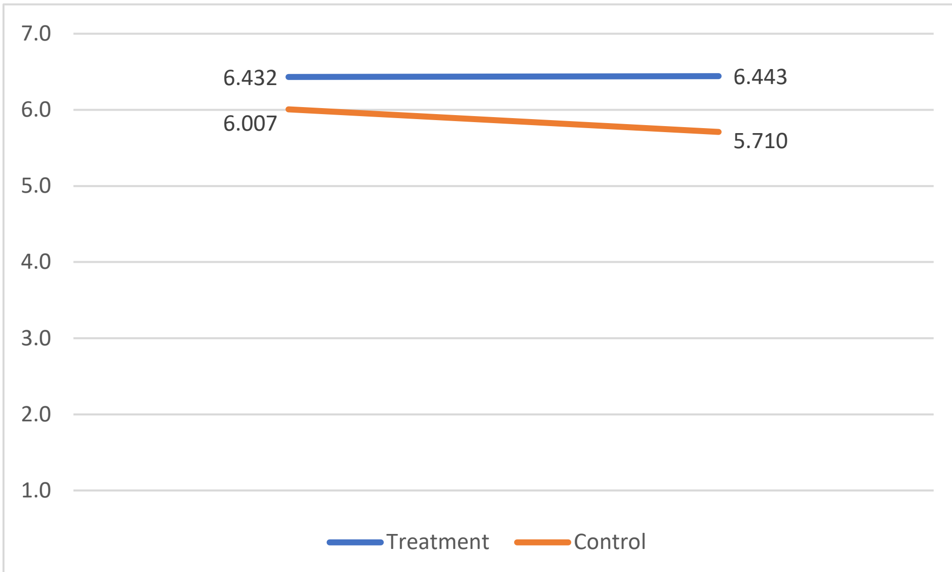


Note. The final adjusted multilevel model for Instructional Support included Baseline Instructional Support and Motivation for MFA PD. The baseline difference for the analytic sample used to estimate the adjusted mean scores below was computed as *g* = 0.41 (Unsatisfied). Correlations between the four CLASS domains range from 0.68 to 0.83 at the pretest and 0.74 to 0.88 at the posttest. Descriptive statistics for the Instructional Support dimensions are shown below (posttest means do not adjust for covariates).

CLASS Instructional Support Dimension	MFA Treatment				BAU Control			
	N	Pretest Mean (SD)	*N*	Posttest Mean (SD)	*N*	Pretest Mean (SD)	*N*	Posttest Mean (SD)
Instructional Learning Formats Learning targets/organization; variety of modalities, strategies, and materials; active facilitation; effective engagement	18	4.56 (1.32)	22	4.62 (0.91)	19	4.25 (1.15)	17	4.11 (0.87)
Content Understanding Depth of understanding; communication of concepts and procedures; background knowledge and misconceptions; transmission of content knowledge and procedures; opportunity for practice of procedures and skills	18	4.06 (1.33)	22	3.69 (1.18)	19	3.68 (1.32)	17	3.03 (0.67)
Analysis and Inquiry Facilitation of higher-order thinking; opportunities for novel application; metacognition	18	1.65 (0.70)	22	2.05 (0.91)	19	1.56 (0.64)	17	1.43 (0.38)
Quality of Feedback Feedback loops; scaffolding; building on student responses; encouragement and affirmation	18	3.50 (1.17)	22	3.70 (1.06)	19	3.17 (1.28)	17	2.79 (0.85)
Instructional Dialogue Cumulative content-driven exchanges; distributed talk; facilitation strategies	18	3.09 (1.46)	22	3.05 (1.07)	19	2.88 (1.41)	17	2.68 (1.04)

Exhibit 17. CLASS Classroom Organization Domain – Year 1 Impact Analyses

Group	Fall 2015 Pretest		Spring 2016 Posttest			Hedges's *g*	*p*-value
	Mean Score	SD	Unadjusted Mean Score	Adjusted Mean Score	SD		
MFA	6.432	0.910	6.472	6.443	0.554	0.775	0.063
BAU	6.007	0.934	5.523	5.710	1.233		

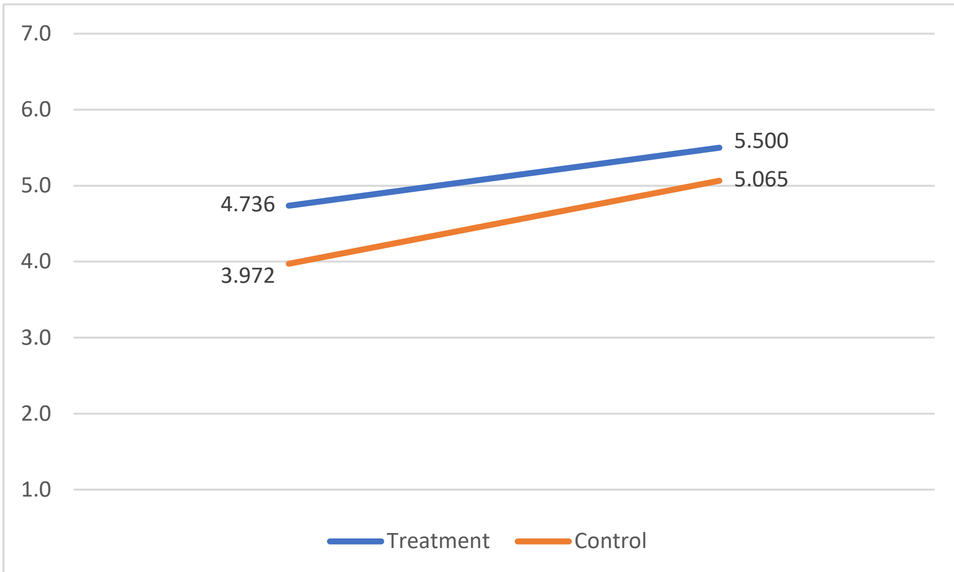


Note. The final adjusted multilevel model for Classroom Organization included Baseline Classroom Organization and Motivation for MFA PD. The baseline difference for the analytic sample used to estimate the adjusted mean scores below was computed as g = 0.45 (Unsatisfied). Correlations between the four CLASS domains range from 0.68 to 0.83 at the pretest and 0.74 to 0.88 at the posttest. Descriptive statistics for the Classroom Organization dimensions are shown below (posttest means do not adjust for covariates).

CLASS Classroom Organization Dimension	MFA Treatment				BAU Control			
	N	Pretest Mean (SD)	*N*	Posttest Mean (SD)	*N*	Pretest Mean (SD)	*N*	Posttest Mean (SD)
Behavior Management Clear expectations; proactive; effective redirection of misbehavior; student behavior	18	6.18 (1.25)	22	6.00 (1.17)	19	6.04 (1.41)	17	5.30 (1.37)
Productivity Maximizing learning time; routines; transitions; preparation	18	6.02 (1.14)	22	6.01 (1.04)	19	5.66 (1.44)	17	5.51 (1.28)
Negative Climate (reverse scored) Absence of negative affect, punitive control, disrespect	18	6.74 (0.54)	22	6.77 (0.48)	19	6.73 (0.53)	17	6.22 (1.14)

Exhibit 18. CLASS Student Engagement Domain – Year 1 Impact Analyses

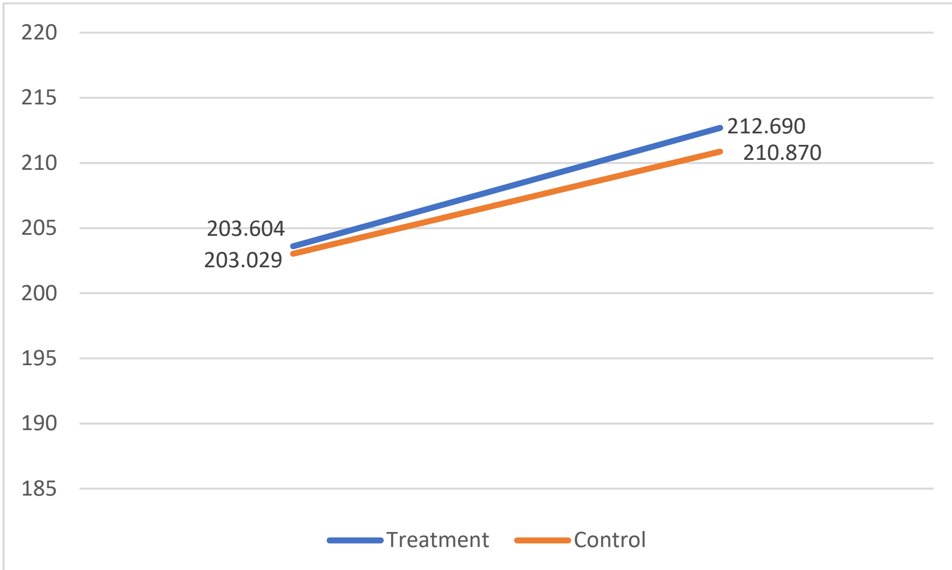
Group	Fall 2015 Pretest		Spring 2016 Posttest			Hedges's *g*	*p*-value
	Mean Score	SD	Unadjusted Mean Score	Adjusted Mean Score	SD		
MFA	4.736	1.171	4.972	5.500	0.670	0.536	0.326
BAU	3.972	1.187	4.426	5.065	0.909		



Note. The final adjusted Ordinary Least Squares model for Student Engagement included baseline Student Engagement and Teacher Race. The baseline difference for the analytic sample used to estimate the adjusted mean scores below was computed as g = 0.66 (Unsatisfied). Correlations between the four CLASS domains range from 0.68 to 0.83 at the pretest and 0.74 to 0.88 at the posttest. The Student Engagement domain has no sub-dimensions, unlike the other three CLASS domains; therefore no additional descriptive statistics are provided.

Exhibit 19. NWEA School-Level – Year 1 Impact Analyses

Group	Spring 2015 Pretest		Spring 2016 Posttest			Hedges's *g*	*p*-value
	Mean Score	SD	Unadjusted Mean Score	Adjusted Mean Score	SD		
MFA	203.604	4.409	212.923	212.690	4.652	0.327	0.086
BAU	203.029	6.553	210.634	210.870	6.115		

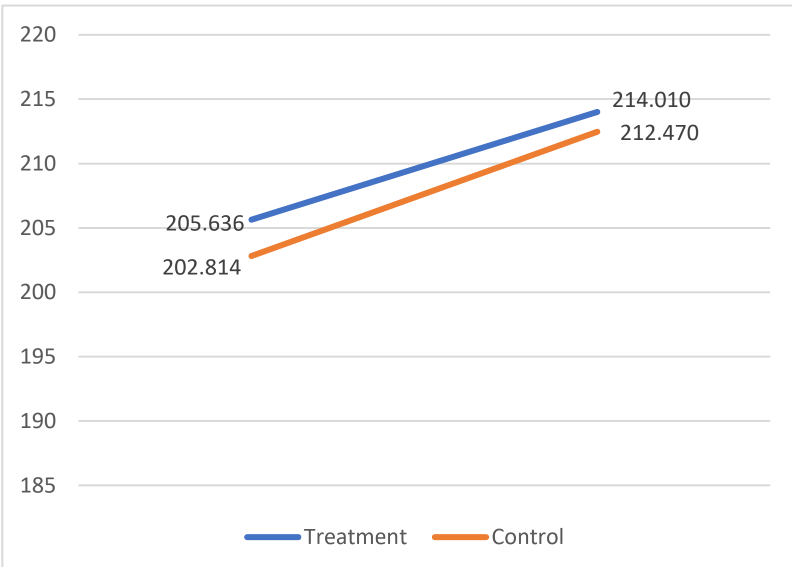


Note. The baseline difference for the analytic sample used to estimate the adjusted mean scores below was computed as g = 0.10 (Statistical Adj.)

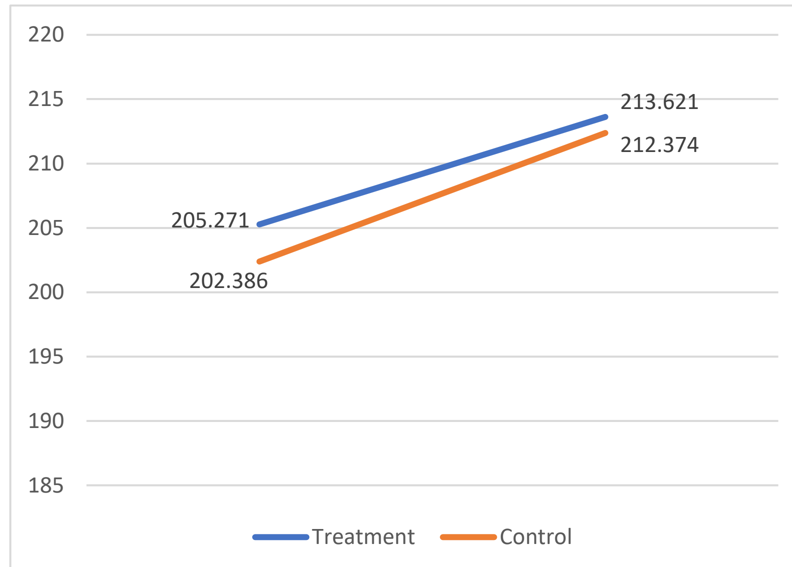
Exhibit 20. NWEA Student-Level – Year 1 Impact Analyses

UNIMPUTED							
	Spring 2015 Pretest		Spring 2016 Posttest				
Group	Mean Score	SD	Unadj. Mean Score	Adj. Mean Score	SD	Hedges's *g*	*p*-value
MFA	205.636	13.495	215.901	214.010	14.514	0.106	0.394
BAU	202.814	14.581	211.760	212.470	14.409		

IMPUTED							
	Spring 2015 Pretest		Spring 2016 Posttest				
Group	Mean Score	SD	Unadj. Mean Score	Adj. Mean Score	SD	Hedges's *g*	*p*-value
MFA	205.271	13.741	215.533	213.621	14.663	0.140	0.187
BAU	202.386	14.705	211.204	212.374	14.864		



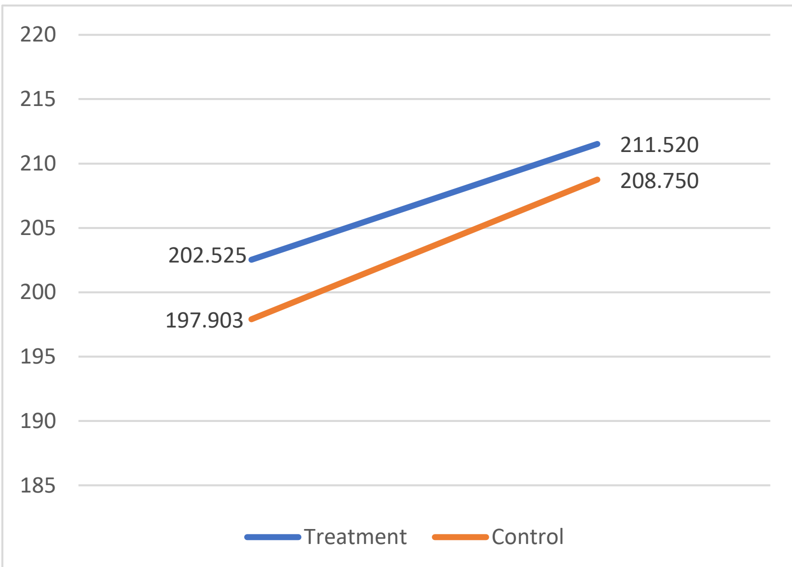
Note. The final model included the pretest, treatment, economically disadvantaged, and ELL, as well as a random intercept and a random ELL slope. The model made use of *n* = 458 BAU students and *n* = 423 MFA students. The baseline difference for the analytic sample used to estimate the adjusted mean scores was computed as *g* = 0.20 (Statistical Adj).



Note. The final model included the pretest, treatment, economically disadvantaged, white, and ELL with a random intercept for each school. Random slopes for each included covariate were explored, as were school-level predictors for all potential covariates, but none improved model fit. The model made use of *n* = 9,660 (483 per imputation) BAU students and *n* = 9,020 (451 per imputation) MFA students. Baseline equivalence analysis yielded a Hedges's *g* = 0.20 (Statistical Adj.).

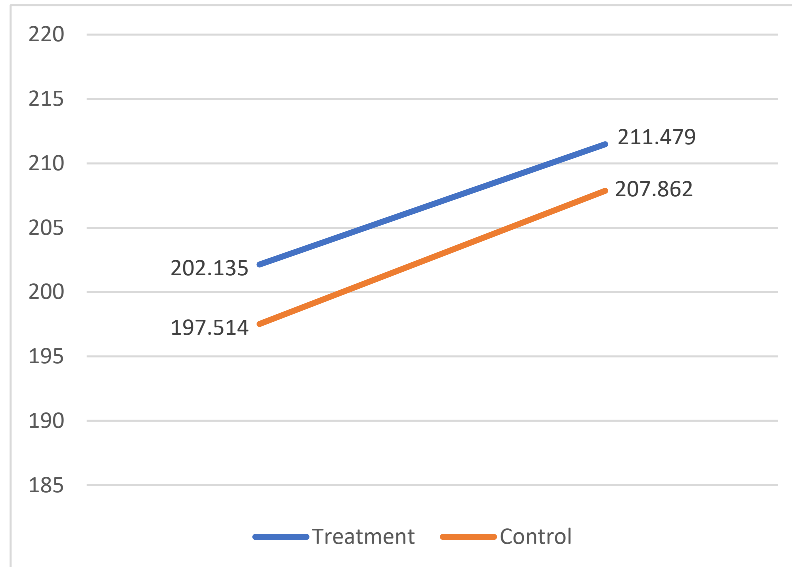
Exhibit 21. Student NWEA Grade 4 – Year 1 Moderator Analyses

UNIMPUTED							
	Spring 2015 Pretest		Spring 2016 Posttest				
Group	Mean Score	SD	Unadj. Mean Score	Adj. Mean Score	SD	Hedges's *g*	*p*-value
MFA	202.525	12.806	214.070	211.520	14.149	0.203	0.159
BAU	197.903	12.827	207.389	208.750	13.053		



Note. The final model included the pretest, treatment, an indicator dummy for grade level = 5, economically disadvantaged, and ELL main effects. The model also included an interaction between the grade dummy and pretest and the grade dummy and treatment indicator, as well as an interaction between ELL and pretest, with random effects for the intercept and the grade dummy slope. Because of the interaction between grade level and treatment, the summary tables are reported for grades 4 and 5 separately. The model made use of n = 247 BAU students and n = 242 MFA students in grade 4. The baseline difference for the analytic sample used to estimate the adjusted mean scores was computed as g = 0.36 (Unsatisfied).

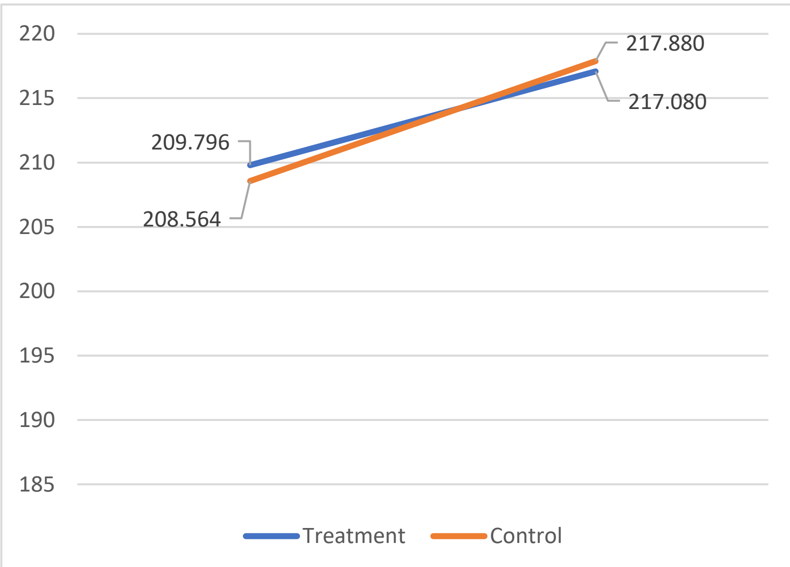
IMPUTED							
	Spring 2015 Pretest		Spring 2016 Posttest				
Group	Mean Score	SD	Unadj. Mean Score	Adj. Mean Score	SD	Hedges's *g*	*p*-value
MFA	202.135	13.205	213.772	211.479	14.299	0.260	0.028
BAU	197.514	12.896	206.965	207.862	13.487		



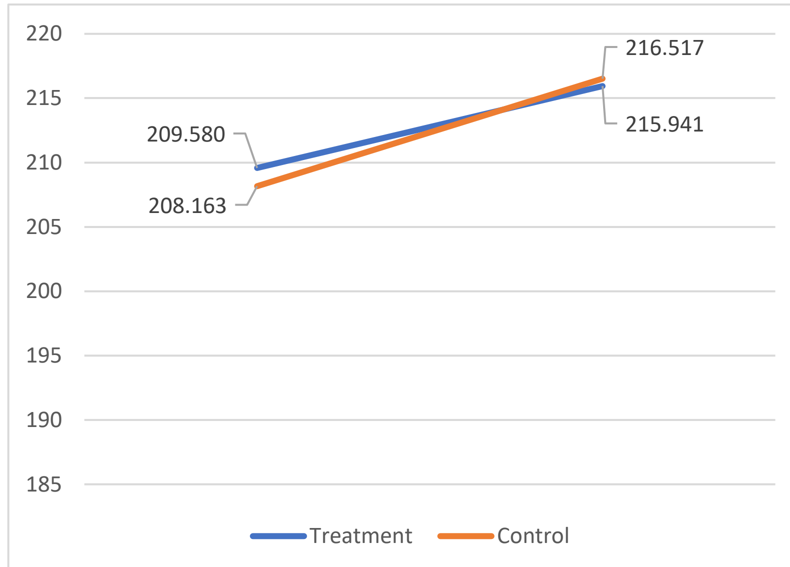
Note. The final model included the pretest, treatment, a grade-level dummy (1 = grade 5), an interaction between grade and treatment, an interaction between grade and pretest, economically disadvantaged, white, ELL, a school-level mean pretest score, and an interaction between pretest and ELL. The final model also included a random intercept for each school. The model made use of n = 5,240 (262 per imputation) BAU students and n = 5,220 (261 per imputation) MFA students in grade 4. Baseline equivalence analysis yielded a Hedges's g = 0.35 (Unsatisfied).

Exhibit 22. Student NWEA Grade 5 – Year 1 Moderator Analyses

UNIMPUTED									IMPUTED								
	Spring 2015 Pretest		Spring 2016 Posttest							Spring 2015 Pretest		Spring 2016 Posttest					
Group	Mean Score	SD	Unadj. Mean Score	Adj. Mean Score	SD	Hedges's *g*	*p*-value		Group	Mean Score	SD	Unadj. Mean Score	Adj. Mean Score	SD	Hedges's *g*	*p*-value	
MFA	209.796	13.304	218.348	217.080	14.672	–0.055	0.649		MFA	209.580	13.289	217.953	215.941	14.813	–0.039	0.739	
BAU	208.564	14.434	216.877	217.880	14.263				BAU	208.163	14.627	216.228	216.517	14.857			

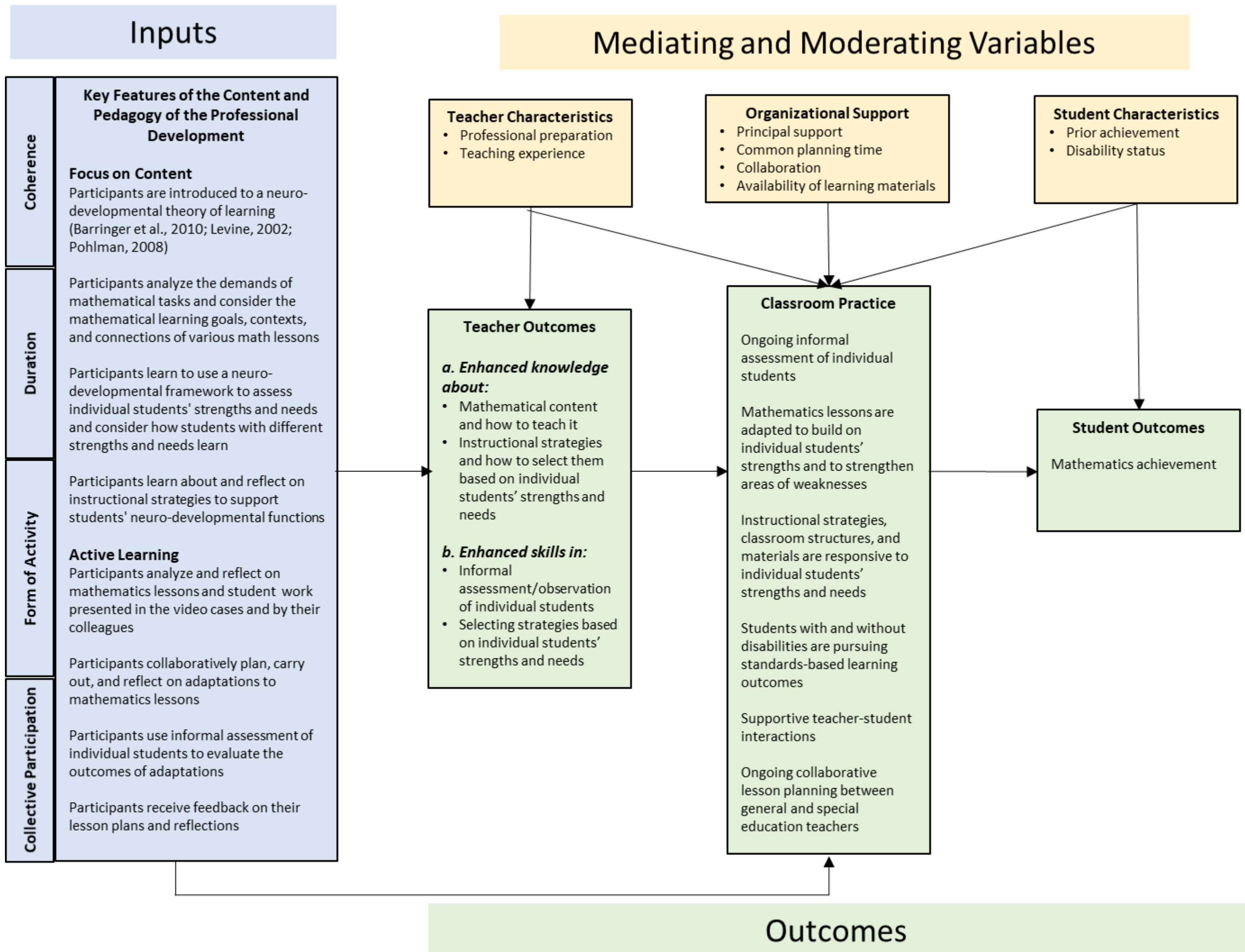


Note. The final model included the pretest, treatment, an indicator dummy for grade level = 5, economically disadvantaged, and ELL main effects. The model also included an interaction between the grade dummy and pretest and the grade dummy and treatment indicator, as well as an interaction between ELL and pretest, with random effects for the intercept and the grade dummy slope. Because of the interaction between grade level and treatment, the summary tables are reported for grades 4 and 5 separately. The model made use of *n* = 211 BAU students and *n* = 181 MFA students in grade 5. The baseline difference for the analytic sample used to estimate the adjusted mean scores was computed as *g* = 0.09 (Statistical Adj.).



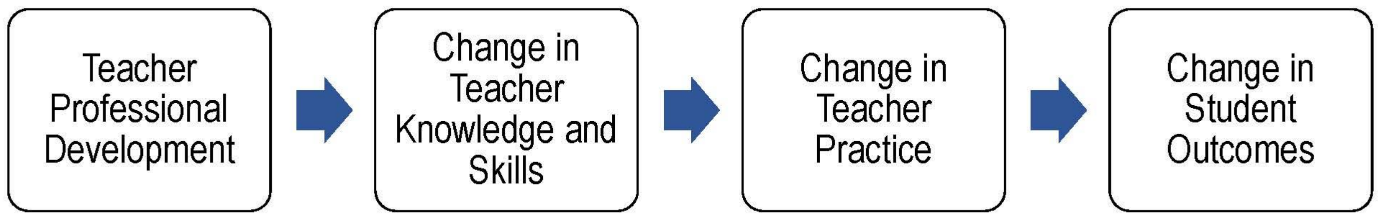
Note. The final model included the pretest, treatment, a grade-level dummy (1 = grade 5), an interaction between grade and treatment, an interaction between grade and pretest, economically disadvantaged, white, ELL, a school-level mean pretest score, and an interaction between pretest and ELL. The final model also included a random intercept for each school. The model made use of *n* = 4,420 (221 per imputation) BAU students and *n* = 3,800 (190 per imputation) MFA students in grade 5. Baseline equivalence analysis yielded a Hedges's *g* = 0.10 (Statistical Adj.).

Appendix A. Math for All Logic Model



Appendix B. Traditional versus dynamic models of teacher professional development theory of change

Traditional Model



Dynamic Model (adapted from Clarke & Hollingsworth, 2002)