Improving Mathematics Achievement for Diverse Learners: The Efficacy of a Professional Development Program for Teachers in Grades K–5

Babette Moeller, Education Development Center; Teresa Duncan & Jason Schoeneberger, Deacon Hill Research Associates; John Hitchcock, Abt Associates; Ellen Meier, Teachers College, Columbia University; Marvin Cohen, Bank Street College of Education

Background

Persistent gaps in mathematics achievement between students (U.S. Department of Education, 2017) calls attention to the fact that teachers often are not well-prepared to implement high-quality instruction with heterogeneous groups of students. While there is great need to improve professional development (PD) efforts, there is little rigorous evidence available to guide this process. Reviews of research on teacher PD (Gersten, Taylor, Keys, Rolfhus, & Newman-Gonchar, 2014; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007) attest to the paucity of relevant studies that link PD to student outcomes. What we do know is that PD can be more effective when it is embedded in a subject area content and focuses on how students learn (Cohen & Hill, 1998; Garet, Porter, Desimone, Birman, & Yoon, 2001; Kennedy, 1998). The Math for All (MFA) PD uses this approach, and this study reports on a test of its impacts on teacher outcomes (knowledge, beliefs, and classroom practice) and student achievement in mathematics.

Setting and Sample

The study took place during the 2015–16 and 2016–17 academic years, involving 32 CPS schools, 96 4th and 5th grade general and special education teachers, and approximately 1,500 4th and 5th grade students. Schools were randomized into MFA and Business as Usual (BAU) conditions. Because of attrition that occurred between first and second years of the study, we focus on results recorded after the first year of the study, where causal validity is strongest (see CONSORT diagram, Exhibit 1). Characteristics of the sample are shown in Exhibits 2-4.

Intervention

Teachers in the intervention group participated in the MFA PD, which 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. MFA consists of five one-day workshops and classroom-based assignments, providing a total of 50 hours of PD over one school year. The program uses video cases and a lesson-study approach to engage teams of general and special education teachers in collaborative lesson planning to make standards-based mathematics lessons accessible to various kinds of learners. Teachers in the BAU comparison group experienced PD offered through their district.

Data Collection and Analysis

The following measures were used:

- Teacher and principal interviews were conducted to gain a deeper understanding of the **implementation** of the intervention.

- **Teacher Mathematical Content Knowledge**. Fourteen questions from the Learning Mathematics for Teaching's item banks (Hill, Rowan & Ball, 2005) 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.
- A teacher survey was administered at the beginning and end of the school year. It included two scales to measure **teachers' Comfort and Preparedness to teach mathematics** to diverse students.
- **Classroom Assessment Scoring System (CLASS) Observation Rubric** measures the quality of teacher-student interactions within four domains: emotional support, classroom organization, instructional support, and student engagement (Pianta, Hamre, & Mintz, 2012). Due to limited resources, we conducted classroom observations of a random subsample of teachers (one grade 4 and one grade 5 teacher in each school).
- **Student Mathematics Achievement** was assessed with the NWEA MAP.

Analyses included multiple imputation, multilevel modeling, and multiple regression. All measures except for the CLASS were converted to Rasch scale scores. Additional details are available in a study brief, online at http://bit.ly/2rkX1qc.

Findings

Implementation data provided evidence of program differentiation between MFA and BAU teachers. While BAU teachers, on average, spent similar amounts of time participating in PD as the MFA teachers, only about half of the BAU teachers who were interviewed reported that they participated in PD focused on mathematics. BAU teachers' interviews also indicate that their mathematics PD did not give them a coherent conceptual model for guiding their work to better understand specific learner needs and to differentiate instruction.

Effect sizes range from 0.106 to 0.982 (Exhibit 5).

- 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 (Exhibit 6).
- We found statistically significant, positive effects of MFA on teachers' reports of **preparedness and comfort** in teaching diverse students (Exhibits 7-8).
- MFA teachers scored higher than teachers in the comparison group on **emotionally supportive classroom practices** (Exhibit 9). This finding is consistent with the MFA PD's emphasis on building teachers' understanding of students' strengths and weaknesses. Although MFA teachers also scored higher in **instructional support, classroom organization, and student engagement**, the classroom observation data were underpowered and these findings did not reach statistical significance (Exhibits 10-12).
- 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 0.327, but not statistically significant (Exhibit 13). 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 non-participating 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 (Exhibit 14).

- 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 MAP scores than students whose teachers were in the BAU group (Exhibit 15). In grade 5, there were very small, non-significant differences between the MFA and comparison groups (Exhibit 16).

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 (Exhibit 17).

References

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.

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.

Gersten, R., Taylor, M. J., Keys, T. D., Rolfhus, E., & Newman-Gonchar, R. (2014). *Summary of research on the effectiveness of math professional development approaches*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

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

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.

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

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP). (2017). *Mathematics assessment data, 4th grade and 8th grade*. Retrieved from https://www.nationsreportcard.gov/ndecore/xplore/nde

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 and parental consent statistics for Math for All 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.

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%

Exhibit 2. Characteristics of MFA schools (randomized sample)

School	Magnet	Title I School	Title 1 School Wide	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. 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%

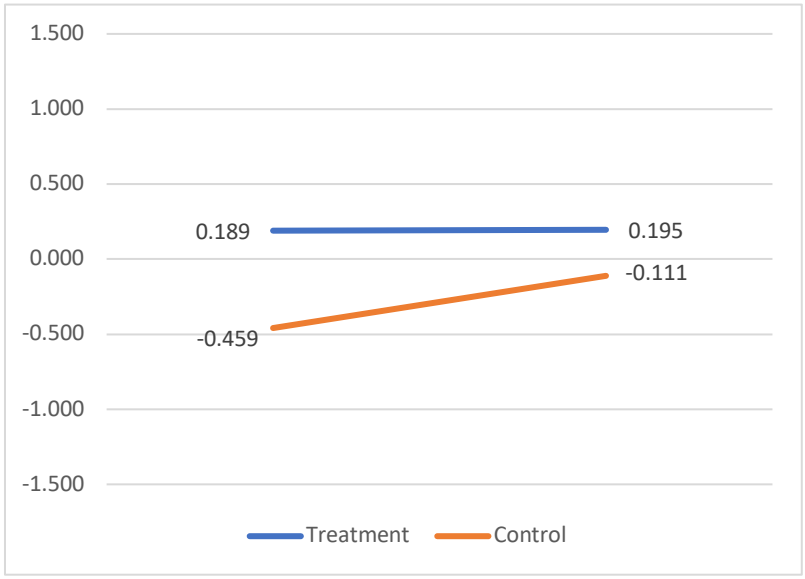
Note. Because parental consent was required to obtain demographic data from the district, we are unable to describe how similar/different the demographics are between the students with parental consent and those without parental consent.

Exhibit 5. Summary of impact analyses results

Outcome at the end of Year 1	Unimputed		Imputed	
	Hedges' g	p-value of treatment indicator	Hedges' g	p-value of treatment indicator
Teacher Knowledge (MKT)	0.432	0.110	0.471	0.074
Teacher Preparedness	0.541	0.013	0.583	0.035
Teacher Comfort	0.666	0.005	0.712	0.009
CLASS Emotional Support	0.982	0.037	N/A	N/A
CLASS Instructional Support	0.690	0.084	N/A	N/A
CLASS Classroom Organization	0.775	0.063	N/A	N/A
CLASS Student Engagement	0.536	0.326	N/A	N/A
NWEA School Level	0.327	0.086	N/A	N/A
NWEA Student Level	0.106	0.394	0.140	0.187
NWEA Student Level - Grade 4	*0.203*	*0.159*	*0.260*	*0.028*
NWEA Student Level - Grade 5	*-0.055*	*0.649*	*-0.039*	*0.739*
Note. Positive values for Hedges' g indicate results that favor the MFA group.				

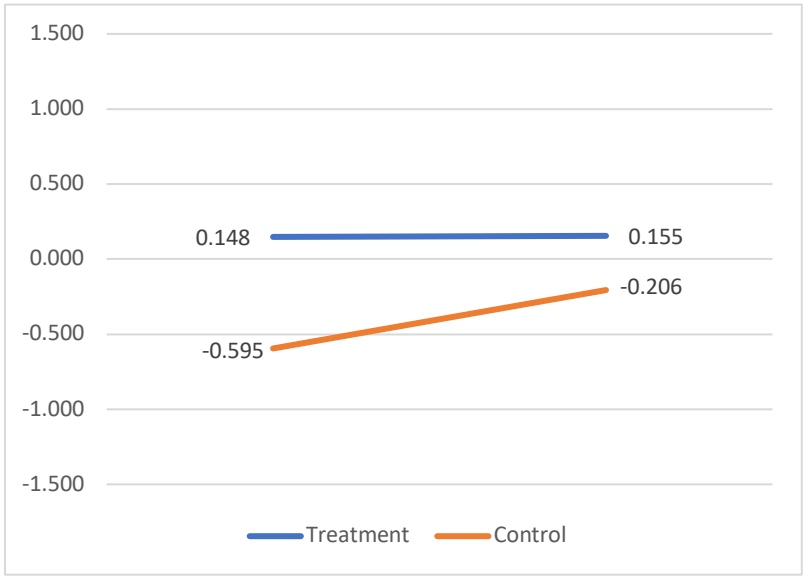
Exhibit 6. Teacher Mathematical Content Knowledge (MKT) – Year 1 Impact Analyses

UNIMPUTED							
	Fall 2015 Pretest		Spring 2016 Posttest				
Group	Mean Score	SD	Unadj. Mean Score	Adj. Mean Score	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 GLH model 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 = .741 (Unsatisfied).

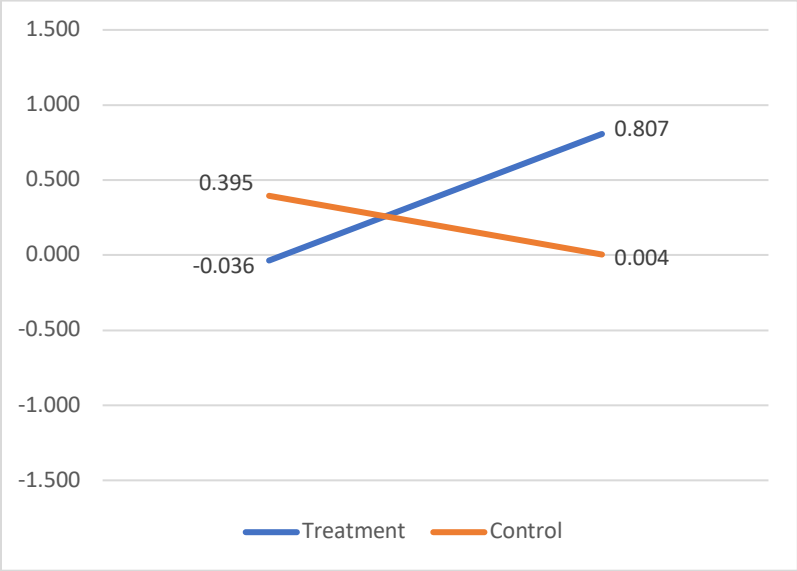
IMPUTED							
	Fall 2015 Pretest		Spring 2016 Posttest				
Group	Mean Score	SD	Unadj. Mean Score	Adj. Mean Score	SD	Hedges' g	p-value
MFA	0.148	0.668	0.232	0.155	0.709	0.471	0.074
BAU	-0.595	1.286	-0.273	-0.206	0.812		



Note. Single-level GLH model 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 = .709 (Unsatisfied).

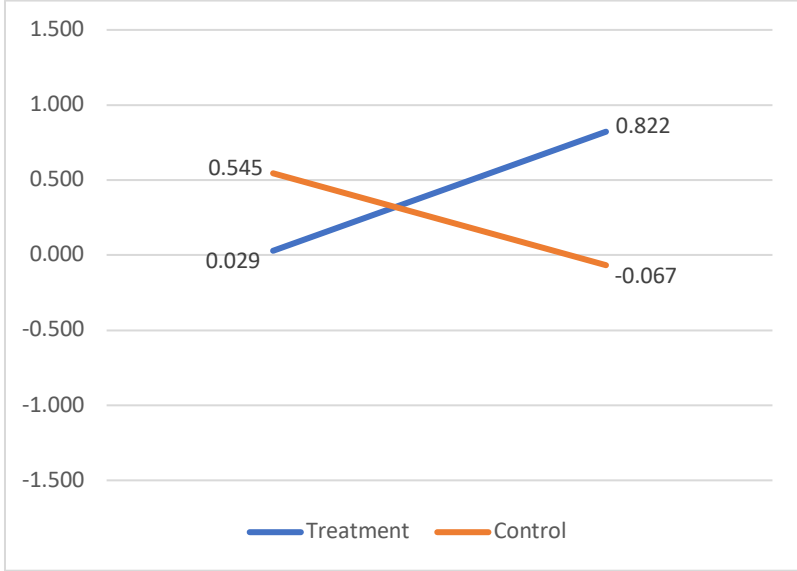
Exhibit 7. Teacher Preparedness in Teaching Students with Disabilities – Year 1 Impact Analyses

UNIMPUTED							
	Fall 2015 Pretest		Spring 2016 Posttest				
Group	Mean Score	SD	Unadj. Mean Score	Adj. Mean Score	SD	Hedges' 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, 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 MFA teachers. The baseline difference for the analytic sample used to estimate the adjusted mean scores below was computed as g = -.36 (Unsatisfied).

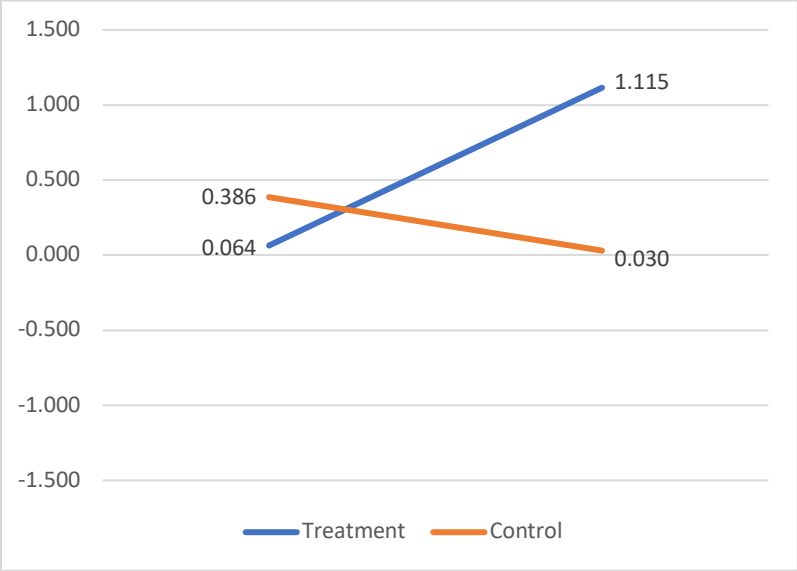
IMPUTED							
	Fall 2015 Pretest		Spring 2016 Posttest				
Group	Mean Score	SD	Unadj. Mean Score	Adj. Mean Score	SD	Hedges' g	p-value
MFA	0.029	1.234	0.811	0.822	1.425	0.583	0.035
BAU	0.545	1.645	-0.087	-0.067	1.603		



Note. Multi-level 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 = -.351 (Unsatisfied; which means the control group had a higher baseline mean).

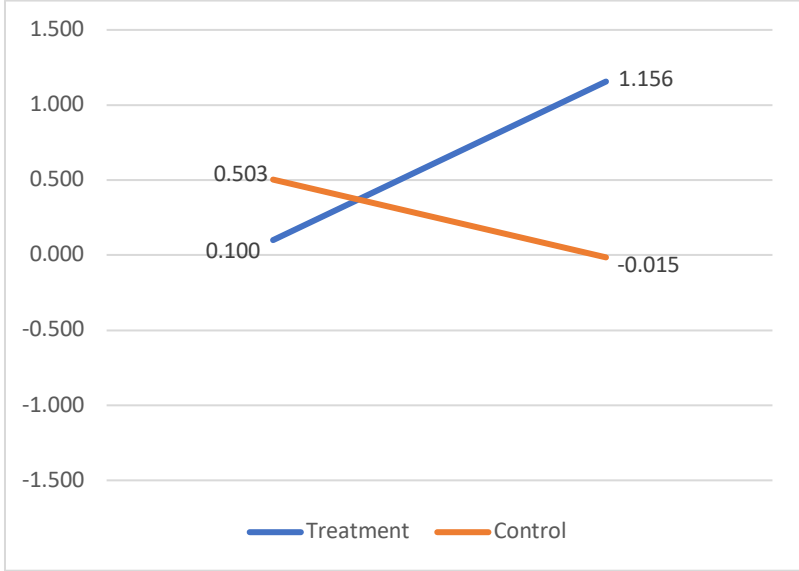
Exhibit 8. Teacher Comfort in Teaching Students with Disabilities – Year 1 Impact Analyses

UNIMPUTED							
	Fall 2015 Pretest		Spring 2016 Posttest				
Group	Mean Score	SD	Unadj. Mean Score	Adj. Mean Score	SD	Hedges' 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 = -.251 (Unsatisfied).

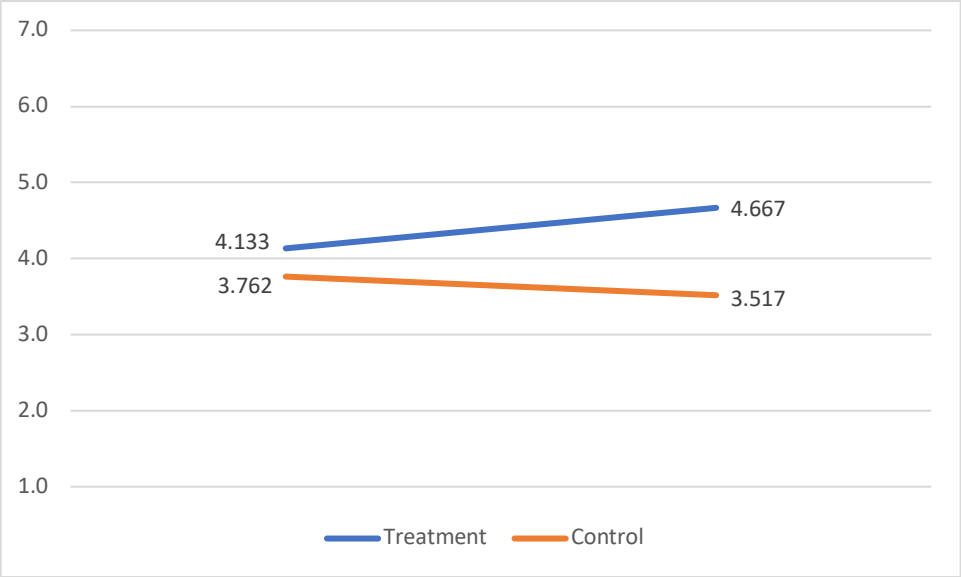
IMPUTED							
	Fall 2015 Pretest		Spring 2016 Posttest				
Group	Mean Score	SD	Unadj. Mean Score	Adj. Mean Score	SD	Hedges' g	p-value
MFA	0.100	1.305	1.143	1.156	1.525	0.712	0.009
BAU	0.503	1.795	-0.037	-0.015	1.738		



Note. Multi-level 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 = -.253 (Unsatisfied; which means the control group had a higher baseline mean).

Exhibit 9. CLASS Emotional Support Domain – Year 1 Impact Analyses

Group	Fall 2015 Pretest		Spring 2016 Posttest			Hedges' 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		

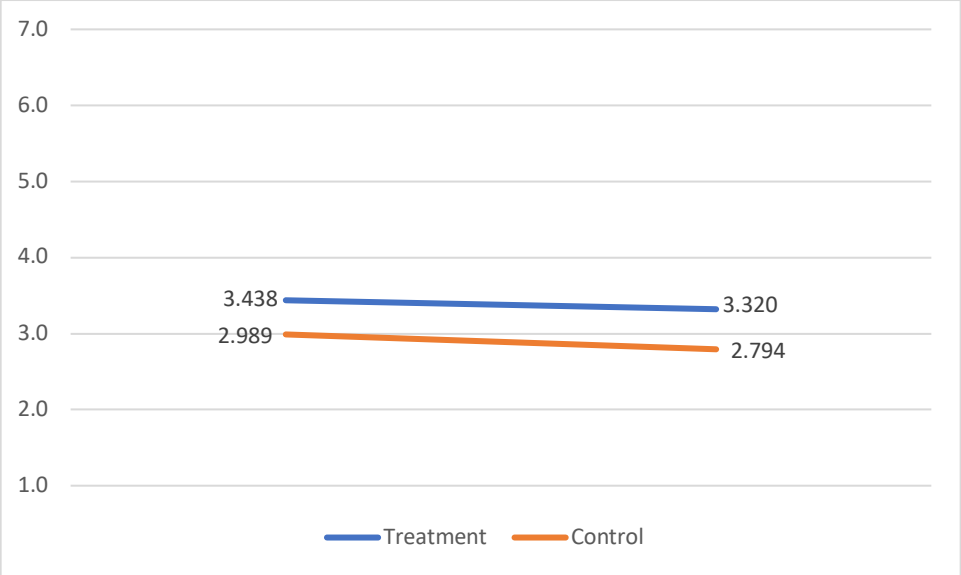


Note. The final adjusted multilevel model for Emotional Support included Baseline Emotional Support, SWD Experience (dichotomized), teacher racial minority indicator, Motivation for MFA PD, and Years Teaching Experience. The baseline difference for the analytic sample used to estimate the adjusted mean scores below was computed as g = .305 (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 = 0.0125. 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 10. CLASS Instructional Support Domain – Year 1 Impact Analyses

Group	Fall 2015 Pretest		Spring 2016 Posttest			Hedges' 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		

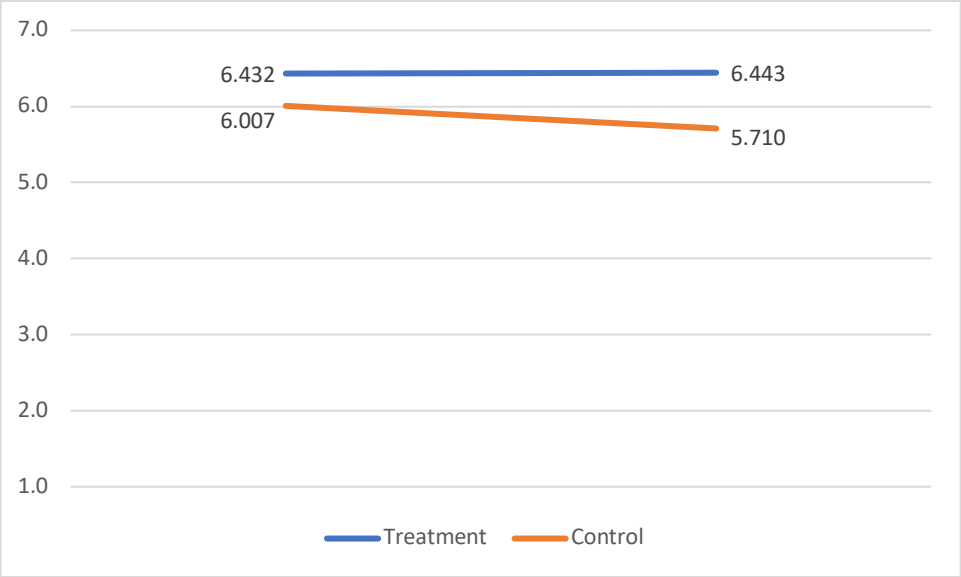


Note. The final adjusted multilevel model for Instructional Support included Baseline Instructional Support and Motivation for MFA PD. The baseline difference for the analytic sample used to estimate the adjusted mean scores below was computed as g = .407 (Unsatisfied). Correlations between the four 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 11. CLASS Classroom Organization Domain – Year 1 Impact Analyses

Group	Fall 2015 Pretest		Spring 2016 Posttest			Hedges' 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		

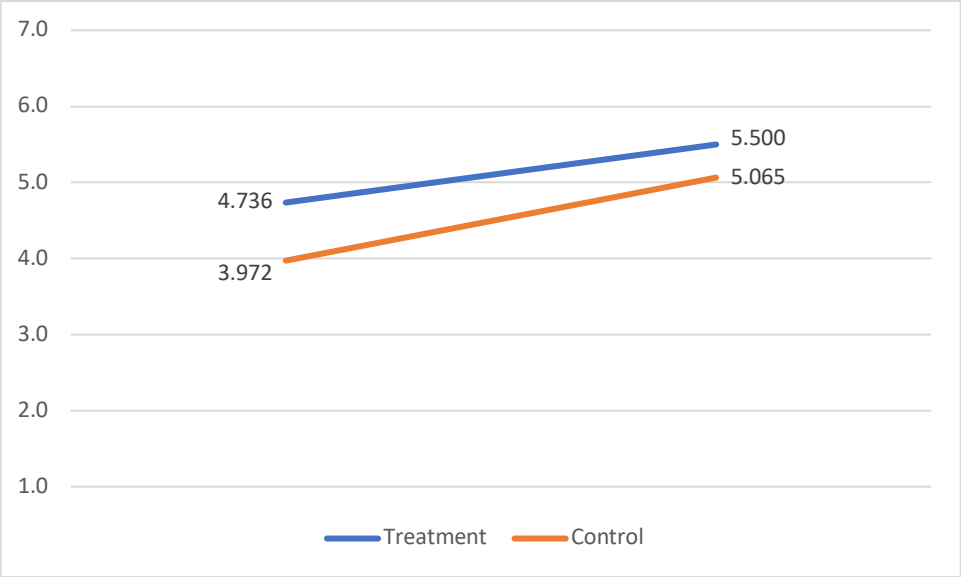


Note. The final adjusted multilevel model for Classroom Organization included Baseline Classroom Organization and Motivation for MFA PD. The baseline difference for the analytic sample used to estimate the adjusted mean scores below was computed as g = .448 (Unsatisfied). Correlations between the four 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 12. CLASS Student Engagement Domain – Year 1 Impact Analyses

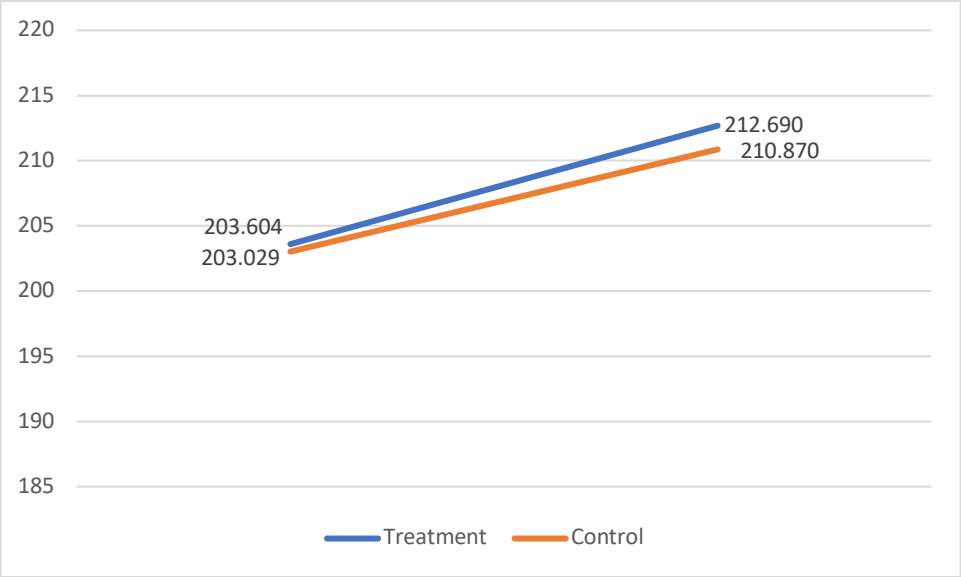
Group	Fall 2015 Pretest		Spring 2016 Posttest			Hedges' g	p-value
	Mean Score	SD	Unadjusted Mean Score	Adjusted Mean Score	SD		
MFA	4.736	1.171	4.972	5.500	0.670	0.536	0.326
BAU	3.972	1.187	4.426	5.065	0.909		



Note. The final adjusted OLS 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 = .663 (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 13. NWEA School-Level – Year 1 Impact Analyses

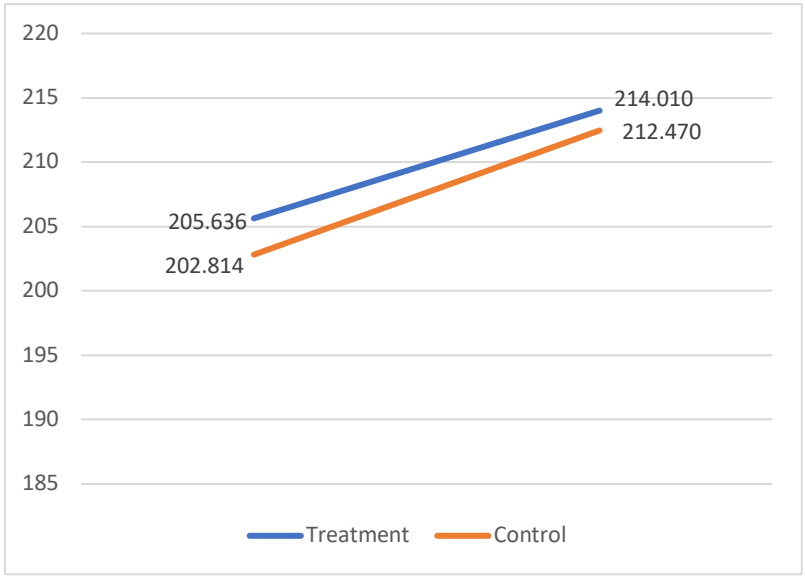
Group	Spring 2015 Pretest		Spring 2016 Posttest			Hedges' g	p-value
	Mean Score	SD	Unadjusted Mean Score	Adjusted Mean Score	SD		
MFA	203.604	4.409	212.923	212.690	4.652	0.327	0.086
BAU	203.029	6.553	210.634	210.870	6.115		



Note. The baseline difference for the analytic sample used to estimate the adjusted mean scores below was computed as g = .100 (Stat. Adjustment)

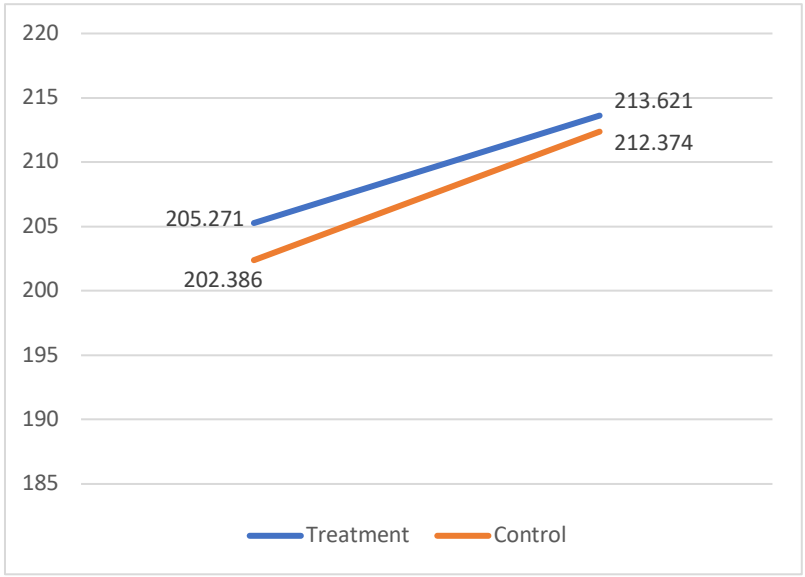
Exhibit 14. NWEA Student-Level – Year 1 Impact Analyses

UNIMPUTED							
	Spring 2015 Pretest		Spring 2016 Posttest				
Group	Mean Score	SD	Unadj. Mean Score	Adj. Mean Score	SD	Hedges' 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		



Note. The final model included the pre-test, 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 = .200 (Statistical Adj).

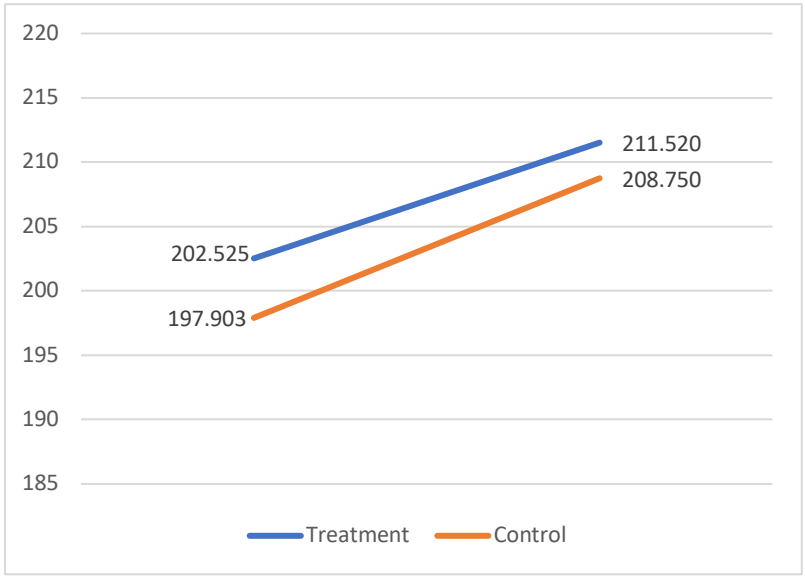
IMPUTED							
	Spring 2015 Pretest		Spring 2016 Posttest				
Group	Mean Score	SD	Unadj. Mean Score	Adj. Mean Score	SD	Hedges' g	p-value
MFA	205.271	13.741	215.533	213.621	14.663	0.140	0.187
BAU	202.386	14.705	211.204	212.374	14.864		



Note. The final model included the pre-test, 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 =9660 (483 per imputation) BAU students and n =9020 (451 per imputation) MFA students. Baseline equivalence analysis yielded a hedges' g = .202 (Statistical Adj.).

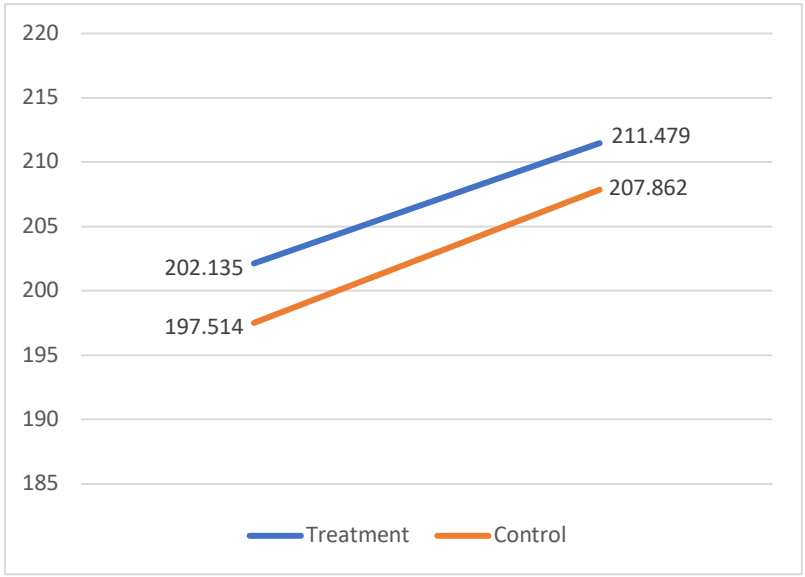
Exhibit 15. Student NWEA Grade 4 – Year 1 Moderator Analyses

UNIMPUTED							
	Spring 2015 Pretest		Spring 2016 Posttest				
Group	Mean Score	SD	Unadj. Mean Score	Adj. Mean Score	SD	Hedges' 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 pre-test, 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 pre-test, 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 = .360 (Unsatisfied).

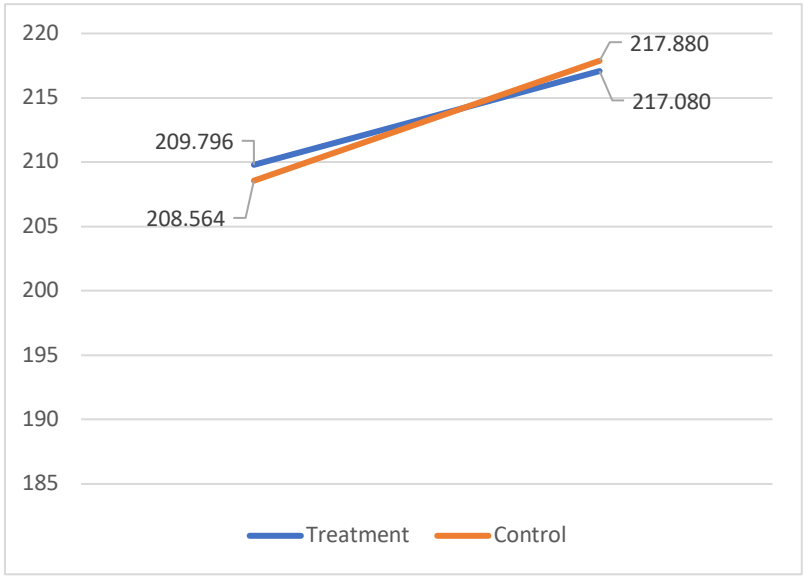
IMPUTED							
	Spring 2015 Pretest		Spring 2016 Posttest				
Group	Mean Score	SD	Unadj. Mean Score	Adj. Mean Score	SD	Hedges' g	p-value
MFA	202.135	13.205	213.772	211.479	14.299	0.260	0.028
BAU	197.514	12.896	206.965	207.862	13.487		



Note. The final model included the pre-test, treatment, a grade level dummy (1=grade 5), an interaction between grade and treatment, an interaction between grade and pre-test, economically disadvantaged, White, ELL, a school-level mean pre-test score and an interaction between pre-test and ELL. The final model also included a random intercept for each school. The model made use of n =5240 (262 per imputation) BAU students and n =5220 (261 per imputation) MFA students in grade 4. Baseline equivalence analysis yielded a Hedges' g = .354 (Unsatisfied).

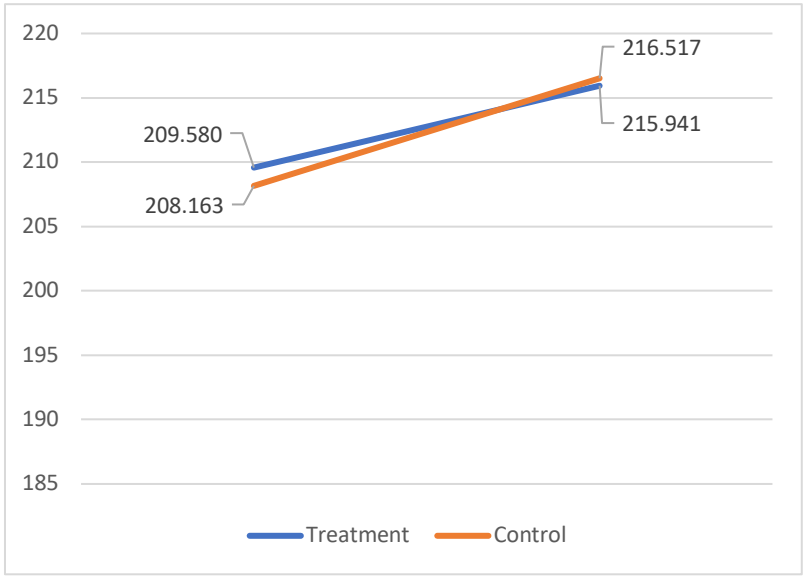
Exhibit 16. Student NWEA Grade 5 – Year 1 Moderator Analyses

UNIMPUTED							
	Spring 2015 Pretest		Spring 2016 Posttest				
Group	Mean Score	SD	Unadj. Mean Score	Adj. Mean Score	SD	Hedges' g	p-value
MFA	209.796	13.304	218.348	217.080	14.672	-0.055	0.649
BAU	208.564	14.434	216.877	217.880	14.263		



Note. The final model included the pre-test, 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 pre-test, 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 = .088 (Stat. Adj).

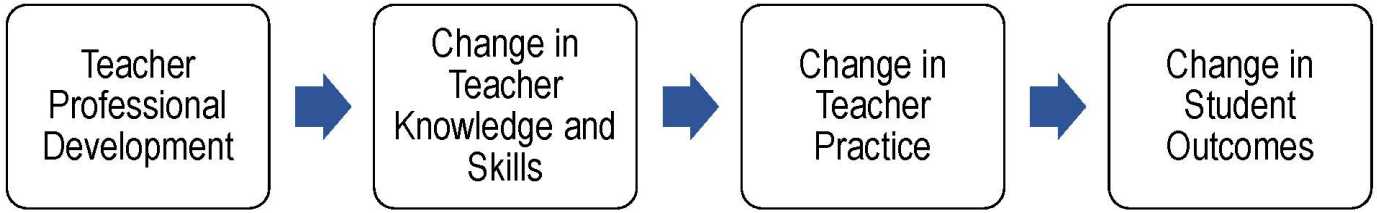
IMPUTED							
	Spring 2015 Pretest		Spring 2016 Posttest				
Group	Mean Score	SD	Unadj. Mean Score	Adj. Mean Score	SD	Hedges' g	p-value
MFA	209.580	13.289	217.953	215.941	14.813	-0.039	0.739
BAU	208.163	14.627	216.228	216.517	14.857		



Note. The final model included the pre-test, treatment, a grade level dummy (1=grade 5), an interaction between grade and treatment, an interaction between grade and pre-test, economically disadvantaged, White, ELL, a school-level mean pre-test score and an interaction between pre-test and ELL. The final model also included a random intercept for each school. The model made use of n =4420 (221 per imputation) BAU students and n =3800 (190 per imputation) MFA students in grade 5. Baseline equivalence analysis yielded a hedges' g = .101 (Statistical Adj.).

Exhibit 17. Traditional versus dynamic models of teacher PD theory of change

Traditional Model



Dynamic Model (adapted from Clarke & Hollingsworth, 2002)