rppl

Annenberg Institute
BROWN UNIVERSITY

Testing the Effects of a Teacher Curriculum Mindset Intervention

January 2025

Hannah Carter
Research Project Manager, Harvard University

Heather Hill
Hazen-Nicoli Professor in Teacher Learning and Practice,
Harvard University | RPPL Senior Researcher

Brendon Krall
Research Project Manager, Annenberg Institute at Brown University

Nathaniel Schwartz
Director of Applied Research and Professor of the Practice,
Annenberg Institute at Brown University | RPPL Senior Researcher

Testing the Effects of a Teacher Curriculum Mindset Intervention

Executive Summary

As districts have taken on large-scale curriculum shifts in recent years – spurred on in part by state legislation – district leaders have increasingly focused on providing professional learning (PL) opportunities aimed at helping teachers engage with and make sense of new instructional materials. These types of curriculum-based professional learning (CBPL) reflect a well-developed consensus from curriculum research that adopting new materials does not necessarily translate into use – and growing evidence that recent teacher uptake of new materials has been slow (Doan et al., 2021; Kaufman et al., 2020).

The Research Partnership for Professional Learning (RPPL) represents a network of professional learning (PL) provider organizations, districts, and researchers working together to answer key questions about how to design more effective PL opportunities for teachers.

Within RPPL's network, which includes professional learning (PL) provider organizations, districts, and researchers, one hypothesis about how to increase the success of curriculum implementation is that CBPL efforts should engage more directly with teacher beliefs and mindsets about new curriculum – and in particular, with teacher concerns about how students might react to new materials that are potentially more demanding than those previously in use. A report from TNTP (2018) highlights the fact that, in many classrooms, students are rarely asked by their teachers to complete tasks at grade level, even when they demonstrate success with the tasks they are assigned.

This research report describes a test of an intervention, conducted as part of an existing research sequence on CBPL, that aimed to provide teachers with opportunities to reflect on their own mindsets around new curriculum. The intervention, which uses a series of written prompts across a defined time period, draws on principles from social psychology, asking teachers to explicitly connect their values and beliefs with the potential promise of high-quality curriculum.

We do not find significant effects of the intervention on most measured outcomes, including teacher survey responses, classroom observations, or student test results. The exception is a marginally significant and moderate-sized (0.30SD) treatment effect on teacher self-reports of using district-adopted curriculum materials. We note several limitations to this study, including both the time period in the school year when the intervention took place and a relatively small sample size, which potentially made it difficult to detect effects.

Background

Curriculum and Curriculum Use

Standards-based reforms, characterized by grade-level specific learning goals meant to guide classroom practice, rely heavily on the presence of well-designed curriculum to support rigor and high-quality instruction (Smith & O'Day, 1990). Such materials guide teachers' daily practice, providing on-grade-level content and suggesting teaching strategies that engage students in thinking and reasoning. Curriculum materials also organize content, ensuring that key material is covered and, often, is arranged into a coherent learning trajectory for students. Recent years have seen both federal and foundation support for the creation and use of 'high-quality' instructional materials (HQIM) that can support standards-based reforms' emphasis on both academic rigor and student engagement in cognitively demanding work.

Despite academics and policymakers regularly encouraging districts and schools to adopt high-quality curriculum materials, progress has been uneven (Polikoff et al., 2020). One reason is that the growth of supplemental and, in recent years, internet-based materials has revolutionized how teachers use curricula. In the 1980s and 90s, teachers largely used textbooks supplied by their districts as primary resources for teaching (Freeman et al., 1983). In a 2020 survey of U.S. K-12 teachers, by comparison, Doan and colleagues (2021) found that in mathematics, 40% of elementary teachers reported regularly using materials they created themselves, with another 18% regularly using supplemental materials produced by their school or district; these numbers rise for middle and high school teachers (see also Blazar et al., 2020). In ELA, a 2015 survey found that most teachers surveyed in three focal states drew on self-created materials, with such materials much more prominent in the middle or high school grades (Opfer et al., 2017).

Teacher-made or supplemental materials, however, may diminish lesson quality and fail to convey disciplinary content in ways consistent with college and career-ready standards. In mathematics, Hill and Lovison (2021), for instance, found that teacher-made and/or "found on the internet" materials correlated with less meaning-oriented and cognitively demanding instruction in a small sample of middle school teachers. Polikoff and Dean (2019) analyzed ELA materials found on teacherspayteachers.com and concluded that these materials lacked alignment to the Common Core State Standards. Summers (2024) found similar results for science materials available on that website. Teachers who cobble together (Kaufman et al., 2020) disparate materials may also sacrifice curriculum coherence and sequences of lessons with embedded student learning trajectories.

Changing Teachers' Mindsets about the Use of High-Quality Instructional Materials

Based on this evidence, states and districts have recently pressed for the implementation of HQIM, encouraging district and teacher adoption of these materials and teacher fidelity to materials once adopted (Kaufman et al., 2020). Yet as recent teacher surveys demonstrate, teacher uptake of HQIM has been slow (Doan et al., 2021; Kaufman et al., 2020). One reason might be teacher mindsets about the role that mandated curricula should play in instruction. Many teachers equate developing or selecting their own lesson materials with professional expertise, and wish to retain control over their curriculum in order to best serve their students. Another reason might be conflicts between teachers' existing beliefs and the approach taken in

standards-aligned curriculum materials. For instance, early scholars found that beliefs about the nature of academic content, beliefs about how students best learn content, and student capacity for challenging content (e.g., Collopy, 2003; Cronin-Jones, 1991; Roehrig & Kruse, 2005) influenced the extent and quality of teachers' use of standards-based curricula. More recently, Charalambous and Philippou (2010), found that teachers' adoption of a reform-oriented set of mathematics materials was affected by their concern about the capacity of their mathematically weaker students to participate in the lessons.

Evidence suggests that teacher beliefs and mindsets remain barriers to teachers' use of standards-aligned curriculum materials today. A recent RAND survey (Prado Tuma et al., 2020), for instance, found that teachers of mathematics gave relatively poor marks to their main set of curriculum materials for meeting the needs of their special education and English-learning students; roughly 30% of responding teachers reported that they view their materials as too challenging for most students. Teachers also reported frequently modifying their materials (Kaufman, 2020), adapting them to students' learning needs, particularly for students below grade level (Prado Tuma et al., 2020). One result of such modification can be the transformation of standards-aligned materials – ones that require student thinking and reasoning – into more conventional instruction through reductions in the cognitive demand of tasks (e.g., Collopy, 2003; Henningsen & Stein, 1997). As well, teachers who believe their district-adopted materials to be too difficult for students may seek to supplement materials with outside resources, including the poor-quality resources described above.

Given these findings, interventions that address teacher mindsets – specifically their beliefs – pose one avenue for encouraging teachers' use of curriculum materials, and in particular their use of curriculum materials that are aligned to rigorous standards for student learning. Gill et al. (2004), for instance, experimentally changed preservice teachers' beliefs about how students learn and the nature of mathematics teaching by having them read expository text on both topics. Charalambous and colleagues (Charalambous et al., 2009) found that a two-year teacher education program featuring content on the nature and history of mathematical thought led to pre-service teachers who were less likely to perceive mathematics as a fixed body of precise truths. And social psychologists have begun recently to use briefer interventions that promote teacher empathy for and affiliation with students in order to change teacher behavior. Gehlbach and colleagues (2016) encouraged high school teachers to think about similarities with their students; students of teachers receiving this treatment perceived better relationships with their teacher and achieved higher grades. Okonofua and colleagues (2016) designed a brief intervention to encourage teachers' empathetic responses to student misbehavior and found that this intervention cut suspension rates in half.

These results motivated the current study, which asks whether a brief intervention targeting teachers' beliefs about students and curriculum materials can change teacher beliefs, enhance teachers' use of curriculum materials and improve instructional quality. We also examine whether the brief intervention improves student outcomes on standardized tests.

Methods

Setting

We partnered with two public school districts and one public charter network in the same Southeastern state to evaluate the effect of the intervention, which was titled Shared Curriculum. The first district was a small city with 11 total schools and a student population of under 10,000. As of 2022, the National Center for Educational Statistics (NCES) records identified 61% of community members as white and 36% of students as having a family income below the poverty level. The charter network consisted of five schools serving under 2,000 students across several small, rural towns. Communities served by this charter were largely African-American (e.g., roughly 55% of the population in one town, 71% for another town), with 13% and 57% of students having a family income below the poverty level in those towns, respectively. The third district was a largely rural town with seven schools, with NCES 2022 data suggesting a largely white (89%) community surrounding the school, and total student enrollment under 5,000. Roughly 20% of students served by these schools came from families with incomes below the poverty level.

All districts had recently adopted HQIM, and contracted with a leading provider of teacher professional development in order to assist teachers' implementation of those materials. All teachers in the study were offered these professional learning opportunities. Teachers were randomly assigned to the intervention, as described below.

Sample

The study involved 155 teachers from the districts and charter network over the course of an academic year. ELA and mathematics teachers formed the bulk of study participants, representing 52% and 22% of the sample respectively (80 ELA teachers and 34 math teachers). However, due to district requests for wider teacher participation, 41 teachers (26%) taught other subjects such as science, social studies, foreign languages, or music. Teachers of all grade levels were represented in the sample, from Pre-K to grade 12. Teachers were evenly distributed among grade bands, with 43 (28%) identified as teachers of students in grades PreK-2, 27 (17%) teaching students in grades 3-5, 30 (19%) teaching students in grades 6-8, and 41 (26%) teaching students in grades 9-12. Fourteen teachers did not have a grade level listed. Twenty-eight teachers (or 18% of the sample) were identified as special education teachers, meaning they were designated by school leaders as the teacher of record for a special education resource class or as a co-teacher for special education students, with the latter typically using a push-in model to support students receiving special education services in general education classes, or a supporting a hybrid approach where students split time between the general education classes and resource classes.

All 155 teachers took a beginning-of-year baseline before the intervention began and were included in randomization. Eight of these teachers exited the study sometime during the implementation year for a variety of reasons, including no longer being with the district (3), no longer being employed (2), medical leave (1), family reasons (1), or no longer being interested in participating in the study (1). However, because they all completed the baseline survey and some had the opportunity to complete the end-of-year post-survey or observations, we kept them in our full sample and several of these teachers remain in some or all of our analyses.

We next describe the three analytic samples used in this report, corresponding to individuals for whom we have complete survey data, individuals for whom we captured a classroom observation, and individuals for whom we were able to obtain student test score data.

Survey Sample

All 155 teachers in the study completed the baseline survey containing items that captured teachers' beliefs about their grade-level standards and students' abilities to meet them, curriculum materials, and how to help students with unfinished learning (a common teacher-cited barrier to using on-grade materials). All teachers were invited to take the end-of-year post-survey. We collected post-survey responses for 119 teachers, or 77% of the full sample. This survey contained items on the topics described above as well as items that asked teachers to reflect on their experience with curriculum implementation. This sample experienced differential attrition between the pre- and post-surveys, with 66 and 53 of treatment and control teachers completing the post-survey, respectively ($p < 0.05$).

Observation Sample

In the spring of 2023, trained raters conducted content-specific classroom observations of participating teachers in both the treatment and control conditions. Of the 155 teachers in the study, a subset of 114 were ELA or math teachers and were eligible for a classroom observation. We collected observation data from 73 teachers (64% of ELA or math teachers). Raters attempted observations for an additional 16 ELA or math teachers (14% of ELA or math teachers) but were unable to complete the observations because of variety of factors, including the teacher or students being absent, disruptions to the school day or regular lesson, or classroom activities that did not lend themselves to being observed (e.g., students reading silently, teacher assigning a project). Classroom observations were not scheduled for 25 ELA or math teachers (22%) for reasons likely due to rater capacity and logistical or scheduling constraints, as the observation window bumped up against state testing and the end of the school year. This sample experienced no differential attrition, with 36 and 37 of treatment and control teachers logging complete observations, respectively ($p > 0.500$).

We collected observation data from 7 teachers of subjects other than ELA or math, but excluded it from analysis due to the lack of a content-specific rubric aligned to the intervention for those subject areas.

Student Administrative Data Sample

Following the academic year, we partnered with districts to obtain student assessment data for teachers who participated in the study. We received student test data for 71 ELA or math teachers (62% of ELA or math teachers in the study, and 46% of the full sample). The assessment data for the 71 teachers represented 2,735 student scores, averaging approximately 38 students per teacher.

Teachers without student test data tended to teach untested grades or subject areas (e.g., calculus). We also received some student achievement data for teachers of other subjects than ELA or math (e.g., science), but excluded it from analysis due to the small sample size. This sample did not experience statistically significant

differential attrition, with 38 and 33 of treatment and control teachers sharing their student outcome data, respectively ($p > 0.500$).

Randomization Design & Procedure

In the fall of 2022, we randomly assigned the 155 participating teachers to either a treatment (i.e., receiving the intervention) or control (i.e., business-as-usual) condition. Randomization ensures that teachers in our volunteer sample were equally likely to be assigned across conditions, making any difference in the observable or unobservable characteristics between groups the result of random chance. To account for differences in contexts, we randomly assigned teachers to a condition within each district, subject area (math or ELA), grade band within math & ELA (Pre-K-8 or 9-12), and "other" subjects or special education status.

We found no statistically significant differences between treatment and control teachers on any of the characteristics we measured at baseline, suggesting that these groups did not differ on observable measures due to random chance in finite samples (Table 1).

Table 1. Baseline Data Balance Table

	Full Sample	Treatment	Control	*p* Value
Grade Level				
PK-2	0.277	0.295	0.260	0.628
3-5	0.174	0.179	0.169	0.862
6-8	0.194	0.192	0.195	0.969
9-12	0.265	0.269	0.260	0.894
Subject				
ELA	0.516	0.513	0.519	0.934
Math	0.219	0.218	0.221	0.966
Other	0.265	0.269	0.260	0.894
Special Education	0.181	0.167	0.195	0.651
Baseline Survey Scales				
Perceptions of Standards	3.439	3.436	3.442	0.972
Expectations of Students	4.537	4.538	4.536	0.981
Student Need for Remediation	2.763	2.854	2.670	0.168

Positive Regard for Curriculum	3.683	3.662	3.704	0.787
Leadership Support for Curriculum	4.139	4.089	4.190	0.425
Curriculum Barriers	3.753	3.688	3.818	0.407
Curriculum Use	3.794	3.833	3.753	0.371
Observations	155	77	78	155

Note: *p<0.1; **p<0.05; ***p<0.01.

For the 77 teachers assigned to receive the intervention, study staff engaged professional learning (PL) organization partners and district and school personnel to encourage teachers to complete the intervention activities. Encouragement occurred via emails to teachers and personal contacts. Teachers were also provided opportunities to complete the intervention activities during their professional learning community (PLC) time within the school day. Teachers in the intervention group completed an introductory module, called "Shared Curriculum Welcome Exercise," between January and February 2023, and completed two follow-up modules during the period of late February to mid-April during the same school year. Teachers in the control group experienced the school year as usual, with no extra professional development provided by either the districts or by the study.

The Shared Curriculum Intervention

The intervention was delivered via Qualtrics to enable the scaling of the intervention while maintaining the precision of its design. Despite using a survey platform, the intervention was designed to encourage teachers' active engagement with the material.

Throughout, the intervention used principles from social psychology to activate teachers' values and connect those values to beliefs and practices that support the use of HQIM. The introductory Shared Curriculum module began by eliciting teachers' written reasons for becoming teachers, then asking respondents to write several sentences about why these reasons were still important to them. This initial prompt sought to elevate teachers' ideal self, one that prioritizes care and connection with students as well as student success. The module then presented teachers with written information about the importance of teacher-student relationships and the importance of teachers listening to students' perspectives. It then pivoted to suggesting that when teachers deviate from the assigned curriculum to make the material less rigorous, students may perceive those actions as meaning that the teacher does not care about students' learning or believe them capable of meeting high standards. The written materials then further cite research and present statements from students themselves to suggest that when teachers support students through difficult material, students can succeed. This pivot connects teachers' values – likely to lie in the area of student success and relationships – to the implementation of rigorous curriculum materials. The intervention positions teachers as experts, asking them to recount successful instances of launching new curriculum materials, including specific ways they overcame challenges. Written text reviews the research literature on how teachers can help students adjust to and master new curriculum materials, and asks teachers to recount an instance in which they have

done the same. The introductory module closes by suggesting to participants that teachers who work together to achieve curriculum mastery will be more successful than those working alone. At many points throughout this exercise, teachers were asked to respond in writing to thirteen separate prompts. The median time to complete the Shared Curriculum Welcome Exercise was roughly 20 minutes.

A first follow-up module, designed for implementation 4-6 weeks after the initial module, used teachers' responses to these prompts, highlighting participating teachers' commitment to seeing all students as capable learners, teachers' commitment to providing grade-level content to students, and teachers' successful use of their curriculum materials. A second follow-up module emphasized teachers working together toward student success, asked teachers to recount successes with their curriculum materials, and again positioned teachers as experts by asking them to write letters to less experienced teachers.

Instrumentation

We collected data from participating teachers and students via a range of instruments. We provide an overview of the instruments here and describe the subsequent measures we constructed from these data below.

Surveys

As noted above, we administered a baseline and post-survey, both with the following constructs. (See Appendix A for items included in each construct).

- *Perceptions of Standards.* We used four items to measure teacher opinions about the grade-level standards of their content area and their assessment of their students' abilities to master them within one school year. A principal components analysis suggested these items form a unidimensional scale, with a Cronbach's alpha of 0.85. We included these items because they might detect impacts of the intervention on teachers' beliefs about the appropriateness of grade-level standards and their students' abilities to meet the demands of rigorous content.

- *Expectations of Students.* We measured teacher expectations of their students with four items assessing the extent to which teachers believed all students, regardless of race and ethnicity, could meet and excel at grade-level appropriate material (see Thomas-Browne et al., 2023). A principal components analysis suggested these items form a unidimensional scale, with a Cronbach's alpha of 0.87. We included these items because they might capture intervention-induced differences in teachers' expectations for students.

- *Student Need for Remediation.* We measured beliefs about remediation with five items assessing teachers' beliefs about remediating students' unfinished learning, for instance, endorsements of teaching basic skills before moving onto more complex tasks and lowering standards for English learners (Hawley et al., n.d.; Teaching Lab, 2024). A principal components analysis suggested these items form a unidimensional scale, with a Cronbach's alpha of 0.70. We included these items because observers of HQIM implementation have identified these beliefs as barriers to the full implementation of HQIM.

- *Positive Regard for Curriculum.* We used a set of five items to measure teachers' beliefs about their district-adopted curriculum materials, including views of the quality of those materials and the appropriateness of those materials for their students. A principal components analysis suggested these items form a unidimensional scale, with a Cronbach's alpha of 0.91. We included these items because they directly tap a belief targeted by the intervention, namely teachers' views of their district-adopted materials.

- *Leadership Support for Curriculum.* We used four items to measure teachers' views of school leader support for curriculum implementation and professional learning around implementation of HQIM, including the extent to which school leaders provide access to time, materials, and resources for implementing learnings from PL and attend the PL with teachers. A principal components analysis suggested these items form a unidimensional scale, with a Cronbach's alpha of 0.86.

- *Curriculum Barriers.* We measured teachers' perceptions of school-level barriers to implementing HQIM on their campus with three items. A principal components analysis suggested these items form a unidimensional scale, with a Cronbach's alpha of 0.74. We included the sets of items on leadership support for and barriers to implementation of HQIM as potential moderators of any treatment impacts; however, due to the small achieved sample, this moderator analysis was not completed.

- *Curriculum Use.* We included a single item to measure teachers' self-reported use of their district-adopted curriculum materials on a 4-point scale ranging from "never" to "every day". This item was intended to capture the intended outcome of the intervention.

The post-survey contained one additional construct:

- *Curriculum Mindset.* We measured teachers' comfort with implementing the curriculum and alignment of that implementation with their beliefs and values through four items. A principal components analysis suggested these items form a unidimensional scale, with a Cronbach's alpha of 0.88. These items corresponded to the mechanism by which the intervention tried to shift curriculum use.

For every construct except for the single-item Curriculum Use construct we used a simple average of the scale items in analysis, as a principal components analysis suggested the items for each construct form a unidimensional scale. Table 2 (below) provides an overview of the summary statistics for each construct.

Table 2. Survey Descriptives

	Baseline Survey					Post-Survey				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Perceptions of Standards	155	3.439	1.013	1	5	119	3.592	.857	1	5
Expectations of Students	155	4.537	.723	1	5	119	4.611	.548	2.5	5
Student Need for Remediation	155	2.763	.828	1	5	119	2.766	.801	1.2	5

Positive Regard for Curriculum	155	3.683	.971	1	5	119	3.736	.975	1	5
Leadership Support for Curriculum	155	4.139	.792	1	5	118	4.034	.854	1	5
Curriculum Barriers	155	3.753	.974	1	5	118	3.893	.873	1	5
Curriculum Use (single item)	155	3.794	.555	1	4	119	3.714	.666	1	4
Curriculum Mindset (post–survey only)	-	-	-	-	-	119	3.637	.811	1	5

Observations

In the spring of 2023, trained raters conducted content-specific classroom observations of participating teachers in both the treatment and control conditions. We selected classroom observation instruments that were developed to capture teacher practices targeted by the intervention, including teaching on-grade material at a high level of student cognitive demand. For ELA, raters used TNTP's Reading Foundational Skills Observation Protocol for observations of early grades (PK-2) ELA instruction (TNTP, n.d.-a) and TNTP's Reading Comprehension Observation Protocol for observations of ELA instruction in grades 3 and above (TNTP, n.d.-b). For observations of mathematics instruction, raters utilized a subset (5 selected rubrics) from the Instructional Practice Research Tool for Mathematics (IPRT-M) (Berlin & Cohen, 2020). The IPRT-M was developed by researchers at the University of Virginia in partnership with Student Achievement Partners to align with constructs captured in the widely used Instructional Practice Guides (Cohen et al., 2020).

We report on each instrument and the number of items here:

- *TNTP Reading Comprehension Observation Protocol.* ELA teachers of students in grades 3 or above were observed using a 4-item observation instrument assessing the extent to which all students were engaged in the work of the lesson; whether the lesson was focused on high-quality texts, questions, and tasks; and whether students demonstrated ownership over doing the work and thinking of the lesson. A principal components analysis suggested these items form a unidimensional scale, with a Cronbach's alpha of 0.91. We used an average of the 4 items in analysis.
- *TNTP Reading Foundational Skills Observation Protocol.* ELA teachers of lower-grades students in PK to 2 were observed using a 5-item observation instrument assessing student engagement, whether the lesson was focused on grade-level reading skills, and the time and quality of the lesson portions devoted to teacher-directed instruction, student practice, and assessment. Even with a small sample size (n = 12), a principal components analysis suggested these items form a unidimensional scale, with a Cronbach's alpha of 0.90. This observation instrument also included a summary rating indicator assessing the extent to which instruction explicitly and systematically provided all students the

opportunity to master foundational skills. We used an average of the 5 items in analysis and analyzed the summary rating separately.

- *Instructional Practice Research Tool for Mathematics (IPRT-M).* Math teachers were observed using 5 selected rubrics from the IPRT-M observation instrument:
    - **Depth of Mathematics** - The teacher made the depth of the mathematics in the lesson explicit through the use of explanations, representations, tasks, and/or examples. The mathematics presented was clear and correct.
    - **Student Representations and Solution Methods** - The teacher strengthened all students' understanding of the content by strategically sharing a variety of students' representations and solution methods.
    - **Prompting Student Thinking** - The teacher posed questions and problems that prompt students to share their developing thinking about the content of the lesson.
    - **Opportunities to Engage in Mathematics** - The teacher provided opportunities for all students to work with and practice mathematics problems, tasks, and exercises.
    - **Student Justifications and Critiques** - Students justified their thinking and critiqued the reasoning of others.

For math observations, classroom observers rated two 15-minute segments of instruction for each teacher. Based on advice from the instrument authors, we averaged the two segment scores for each teacher and then analyzed each item separately rather than trying to form a single scale. Finally, for all subject areas, we asked each observer to rate the approximate percentage of the lesson for which the teacher or students were engaging with the assigned curriculum (from 0 to 100 percent). This item was intended to capture the intended outcome of the intervention.

Given the variation in observation instruments and small sample size, we generated standardized scores. The observation scores were standardized within instrument by subtracting the mean and dividing by the standard deviation, creating a standardized score that facilitates comparison across observation instruments. The resulting scores were non-normally distributed, following a more uniform distribution.

Table 3. Observation Scale Descriptives by Protocol

	N	Mean	SD	Min	Max
Standardized observation score	73	0	.986	-2.252	1.540
TNTP Reading Comprehension Observation Protocol Average	46	2.554	.987	1	4
TNTP Reading Foundational Skills Observation Protocol Average	12	2.383	.711	1.2	3.4
TNTP Reading Foundational Skills Observation Protocol – Summary Rating	12	2.333	.985	1	4

IPRT-M: Depth of Mathematics	15	2.2	.802	1	3
IPRT-M: Student Representations and Solution Methods	15	1.267	.562	1	2.5
IPRT-M: Prompting Student Thinking	15	1.9	.431	1	2.5
IPRT-M: Opportunities to Engage in Mathematics	15	3.167	.957	2	4
IPRT-M: Student Justifications and Critiques	15	1.233	.320	1	2
Percent curriculum use question	71	81.676	27.419	0	100

Student Test Score Data

Following the academic year, we partnered with districts to obtain student assessment data for teachers who participated in the study. As noted above, the intervention was delivered less than four months before the start of state testing, reducing the possibility for impacts on this outcome.

The primary outcome of interest is student performance as measured by end-of-year test scores for the 2022-2023 school year. Scaled test scores were shared for three core subjects: mathematics, English language arts (ELA), and, in one district, science. Districts shared different test results depending on their grade levels served:

- One district and the charter network provided data from the End of Year State Assessment.
- Due to differences in grade levels (early elementary school), one district shared scores from the Star Literacy Assessment.

Given the variation in assessment tools, we standardized the scores to enhance comparability across different districts, tests, and grade levels. The scores were standardized within each district, grade level, and test by subtracting the mean and dividing by the standard deviation, creating a rescaled score that facilitates direct comparison across subjects and districts. We did not receive prior year test scores for most students, so we did not use prior scores in our analysis.

Table 4. Student Scaled Score Descriptives by Subject Area

	N	Mean	SD	Min	Max
Math	1202	419.19467	7.7916479	403	445
ELA PK-2	226	792.57965	71.380394	621	1084

ELA 3+	1309	417.87155	8.3090839	401	445

Implementation Fidelity

During the intervention year, we tracked the extent to which teachers in the treatment group participated in the intervention's introductory module and follow-up exercises. Table 5 (below) includes a breakdown of the intervention timing and completion rates by group. Relatively high module completion rates were achieved in the treatment group, with nearly 3/4 of teachers completing all 3 modules. There was a small amount of contamination, as 4 teachers in the control group (5%) completed the first follow-up exercise. In all, in the treatment group, only 1 teacher (1%) completed 0 modules, 2 teachers (3%) completed 1 module, 17 teachers (22%) completed 2 modules, and 57 teachers (74%) completed all 3 modules. In the control group, 74 teachers (95%) did not complete any modules and 4 teachers (5%) completed 1 module.

Table 5. Timing of Intervention and Completion Rates

	Dates Administered	Completion (Treatment Group)	Completion (Control Group)
Shared Curriculum Welcome Exercise	1/12/23 - 2/15/23	61 / 77 (79.2%)	0 / 78 (0%)
Follow-Up Exercise 1	2/22/23 - 4/11/23 (half (39) by 3/1 and majority (68) by 3/8)	74 / 77 (96.1%)	4 / 78 (5.1%)
Follow-Up Exercise 2	4/4/23 - 4/20/23 (majority (67) by 4/14)	72 / 77 (93.5%)	0 / 78 (0%)

Analytic Approach

We estimate treatment effects on teacher and student outcomes using ordinary least squares regressions. We begin by fitting the following ordinary least model for teacher-level outcomes, where Y represents a given outcome for teacher j:

$$Y_j = \beta Treat_j + \gamma X_j + \pi_k + \varrho_j + \alpha_j + \theta_j + \varepsilon \qquad (1)$$

Here, β captures the intent-to-treat treatment effect of offering teachers the intervention. Xj captures performance on our baseline survey, if applicable to the analysis (e.g., post-survey and observation outcomes). In all models, we include fixed effects for districts π, subject ϱ, grade-level α, and special education teaching status θ. We estimate robust standard errors clustering at the teacher level across all models for teacher-level outcomes.

We analyze student achievement outcomes using an augmented version of Equation 1, excluding any controls.

$$A_{ij} = \beta Treat_j + \pi_k + \varrho_j + \alpha_j + + \theta_j + \varepsilon \qquad (2)$$

Here A_{ij} represents student achievement on the summative achievement test, and π, ϱ, α, and θ represent the fixed effects specified above. Again, we estimate robust standard errors clustered at the teacher level.

Results

Survey Data

Table 6 indicates that the treatment intervention did not produce a statistically significant effect on most teachers' self-reported outcomes as measured by the seven survey scales. On the single item that asked for teacher self-reports of using district-adopted curriculum materials, we found a marginally significant and moderate-sized (0.30SD) treatment effect ($p < 0.10$). These findings suggest that, within the scope of this study and the given sample, the treatment did not lead to a measurable shift in teacher perceptions as measured by a post-intervention survey.

Table 6. Effects of Treatment on Teacher Survey Responses

	Teacher EOY Survey Responses							
	All Scales							
Variables	(1) Perceptions of Standards	(2) Expectations of Students	(3) Student Need for Remediation	(4) Positive Regard for Curriculum	(5) Leadership Support for Curriculum	(6) Curriculum Barriers	(7) Curriculum Use	(8) Curriculum Mindset
Treatment Impact	0.047	0.032	-0.194	-0.108	0.017	0.045	0.175*	0.073
	(.142)	(.114)	(.147)	(.129)	(.133)	(.148)	(.091)	(.161)
Constant	1.852***	2.776***	1.543**	1.172***	1.839***	1.933***	0.843*	3.924***
	(.397)	(.738)	(.504)	(.429)	(.469)	(.538)	(.412)	(.266)
Observations	111	111	111	110	110	111	111	111
R^2	0.4808	0.2507	0.5228	0.6498	0.6326	0.4000	0.5449	0.3483
Adjusted R^2	0.3583	0.0739	0.4102	0.5671	0.5449	0.2568	0.4376	0.2047

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors are robust to heteroskedasticity. All models include controls for baseline survey results, when available, and fixed effects for district, subject, grade-level, and special education teaching status.

Results were consistent across an unconditional model, a model that controlled for baseline survey results alone, a model with only district fixed effects, and a model with fixed effects for district, subject, grade-level, and special education teaching status (but no controls for baseline survey results).

Observation Data

Tables 7 and 8 indicate that the treatment intervention did not produce a statistically significant effect on end-of-year observation scores for treated teachers – either in their standardized overall observation rating or the percentage of time the classroom observer witnessed the teacher or students using the assigned curriculum. These findings suggest that, within the scope of this study and the given sample, the treatment did not lead to measurable differences in teacher observation scores between teachers who received the intervention and those who did not. Results were consistent across an unconditional model, and a model that controlled for both baseline survey scale scores and all relevant fixed effects stemming from randomization. We additionally conducted a sensitivity analysis to determine whether there were differences in scores between treated and control teachers by observation instrument or grade level and did not detect any significant differences (see Appendix C).

Table 7. Effects of Treatment on Teacher Observation Scores – Standardized Across Instruments

	Teacher EOY Observation Scores Standardized	
Variables	(1) Unconditional	(2) Fixed Effects
Treatment Impact	0.080	0.220
	(0.232)	(0.250)
Constant	-0.039	1.427
	(0.174)	(1.158)
Baseline survey controls		x
District FE		x
Grade-Level FE		x
Subject FE		x
SPED FE		x
Observations	73	72
R^2	0.0017	0.3661
Adjusted R^2		0.1175

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors are robust to heteroskedasticity.

Table 8. Effects of Treatment on Teacher Observation Scores – Curriculum Use

	Teacher EOY Observation Scores Curriculum Use Item	
Variables	(1) Unconditional	(2) Fixed Effects
Treatment Impact	1.221	6.125
	(6.584)	(6.898)
Constant	81.057	220.428
	(5.383)	(59.336)
Baseline survey controls		x
District FE		x
Grade-Level FE		x
Subject FE		x
SPED FE		x
Observations	71	70
R^2	0.0005	0.3929
Adjusted R^2		0.1273

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors are robust to heteroskedasticity.

Student Data

Table 9 indicates that the treatment intervention did not produce a statistically significant effect on student outcomes for treated teachers. This was true for both an unconditional model, and a model with fixed effects corresponding to randomization blocks. This finding suggests that, within the scope of this study and the given sample, the treatment did not lead to measurable improvements in student performance. To assess the robustness of our results, we conducted a series of alternative model specifications, all of which supported our primary finding (see Appendix D).

Table 9. Effects of Treatment on Student Performance

	Student EOY Scores	
Variables	(1) Unconditional	(2) Fixed Effects
Treatment Impact	-0.076	-0.077
	(0.125)	(0.144)

Constant	0.072	0.073
	(0.074)	(0.077)
District FE		x
Grade-Level FE		x
Subject FE		x
SPED FE		x
Observations	2735	2735
R^2	0.001	0.066
Adjusted R^2		0.006

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors are clustered at the teacher level.

Discussion

This report analyzes the effects of an intervention designed to support effective teacher uptake of new curriculum by offering teachers additional opportunities to reflect on the ways this curriculum might align with their core values. Similar types of interventions have had large effects on other aspects of teacher practice such as the choices they make around student discipline (Okonofua et al., 2016, Okonofua et al., 2022), but to our knowledge, this strategy has never been tested as a way to shift teacher mindsets around curriculum use.

In this case, the intervention appears to have had little effect on most outcomes, including teachers' observed practice and their students' outcomes on standardized tests. The exception was a marginally significant and moderate-sized (0.30SD) shift on teacher self-reports of using district-adopted curriculum materials.

There are several potential explanations for these results. The amount of time teachers spent in reflection exercises was quite short and disconnected from other types of CBPL happening simultaneously. There are many well-documented reasons that teachers do not make use of new instructional materials – or that they make surface use of new materials without substantially shifting core practice – and it is possible that the complexity of the issue made the intervention studied here too weak to make a measurable difference.

At the same time, there were also several limitations to the study design that might have weakened the potential impact of the intervention. The sample size was lower than our power calculations suggested might be necessary, weakening our ability to detect effects. But, most importantly, due to the slow assignment of teachers to coaches and PL facilitators, the intervention didn't launch until January rather than starting along with the school year. Since teachers had been using the new curriculum for a full semester by that time, they had likely already formed opinions about that curriculum and their relationship to the curriculum that were more difficult to shift. A related concern about the late start was that it meant that teachers had far less time to enact changes in their work either before our observations occurred (only several weeks after the completion of the intervention) or before student testing took place.

This study represented an early attempt with the RPPL network to build randomized interventions that respond to hypotheses co-developed across PL organizations and PL researchers. While it was disappointing not to see bigger effects from the intervention, we see this study as an important step toward more frequent, quick turnaround PL interventions that allow the network to learn together about what it will take to improve PL at scale (Alicea et al., 2024).

References

Alicea, S., Fu, C., Johnson, S., Papay, J., Schwartz, N. (2024). A new model for supporting research-aligned teacher development: Emerging insights from the research partnership for professional learning. https://rpplpartnership.org/external-resource/a-new-model-for-supporting-research-aligned-teacher-development-emerging-insights-from-the-research-partnership-for-professional-learning/

Berlin, R. & Cohen, J. (2020). The convergence of emotionally supportive learning environments and college and career ready mathematical engagement in upper elementary classrooms. AERA-Open, 6(3), 1-20.

Blazar, D., Heller, B., Kane, T. J., Polikoff, M., Staiger, D. O., Carrell, S., Goldhaber, D., Harris, D. N., Hitch, R., Holden, K. L., & Kurlaender, M. (2020). Curriculum reform in the Common Core era: Evaluating elementary math textbooks across six US states. Journal of Policy Analysis and Management, 39(4), 966-1019.

Charalambous, C. Y., & Philippou, G. N. (2010). Teachers' concerns and efficacy beliefs about implementing a mathematics curriculum reform: Integrating two lines of inquiry. Educational Studies in Mathematics, 75(1), 1-21.

Charalambous, C. Y., Panaoura, A., & Philippou, G. (2009). Using the history of mathematics to induce changes in preservice teachers' beliefs and attitudes: Insights from evaluating a teacher education program. Educational Studies in Mathematics, 71, 161-180.

Cohen, J., Hutt, E., Berlin, R., & Wiseman, E. (2020). The change we cannot see: Instructional quality and classroom observation in the era of common core. Educational Policy. https://doi.org/10.1177/0895904820951114

Collopy, R. (2003). Curriculum materials as a professional development tool: How a mathematics textbook affected two teachers' learning. The Elementary School Journal, 103(3), 287-311.

Cronin-Jones, L. L. (1991). Science teacher beliefs and their influence on curriculum implementation: Two case studies. Journal of research in science teaching, 28(3), 235-250.

Doan, S., Fernandez, M. P., Grant, D., Kaufman, J. H., Setodji, C. M., Snoke, J., Strawn, M., & Young, C. J. (2021). American Instructional Resources Surveys: 2021 technical documentation and survey results. RAND Corporation. https://www.rand.org/pubs/research_reports/RRA134-10.html.

Freeman, D. J., Kuhs, T. M., Porter, A. C., Floden, R. E., Schmidt, W. H., & Schwille, J. R. (1983). Do textbooks and tests define a national curriculum in elementary school mathematics? The Elementary School Journal, 83(5), 501-513.

Gehlbach, H., Brinkworth, M. E., King, A. M., Hsu, L. M., McIntyre, J., & Rogers, T. (2016). Creating birds of similar feathers: Leveraging similarity to improve teacher–student relationships and academic achievement. Journal of Educational Psychology, 108(3), 342.

Gill, M. G., Ashton, P. T., & Algina, J. (2004). Changing preservice teachers' epistemological beliefs about teaching and learning in mathematics: An intervention study. Contemporary Educational Psychology, 29(2), 164-185.

Hawley, W., Irvine, J. J., & Landa, M. (n.d.). Common beliefs survey. Learning for Justice. https://www.learningforjustice.org/sites/default/files/general/Common%20Beliefs%20Survey%20New_1.pdf

Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. Journal for Research in Mathematics Education, 28(5), 524-549.

Hill, H. C., & Lovison, V. (2021). US middle school mathematics instruction, 2016. EdWorkingPapers, Annenberg Institute at Brown University.

Kaufman, J. H., Doan, S., Prado Tuma, A., Woo, A., Henry, D., & Lawrence, R. A. (2020). How instructional materials are used and supported in U.S. K–12 classrooms: Findings from the 2019 American Instructional Resources Survey. RAND Corporation. https://www.rand.org/pubs/research_reports/RRA134-1.html

Okonofua, J. A., Paunesku, D., & Walton, G. M. (2016). Brief intervention to encourage empathic discipline cuts suspension rates in half among adolescents. Proceedings of the National Academy of Sciences, 113(19), 5221-5226.

Okonofua, J. A., Goyer, J. P., Lindsay, C. A., Haugabrook, J., & Walton, G. M. (2022). A scalable empathic-mindset intervention reduces group disparities in school suspensions. Science advances, 8(12).

Opfer, V. D., Kaufman, J. H., & Thompson, L. E. (2017). Implementation of K–12 state standards for mathematics and English language arts and literacy: Findings from the American Teacher Panel. RAND Corporation. https://www.rand.org/pubs/research_reports/RR1529-1.html

Polikoff, M. & Dean, J. (2019). The supplemental curriculum bazaar: Is what's online any good? Thomas B. Fordham Institute. https://fordhaminstitute.org/national/research/supplemental-curriculum-bazaar

Polikoff, M., Wang, E. L., Haderlein, S. K., Kaufman, J. H., Woo, A., Silver, D., & Opfer, V. D. (2020). Exploring coherence in English language arts instructional systems in the Common Core era. RAND Corporation. https://www.rand.org/content/dam/rand/pubs/research_reports/RRA200/RRA279-1/RAND_RRA279-1.pdf

Prado Tuma, A., Doan, S., Lawrence, R. A., Henry, D., Kaufman, J. H., Setodji, C. M., Grant, D., & Young, C. J. (2020). American Instructional Resources Survey: 2019 technical documentation and survey results. RAND Corporation. https://www.rand.org/pubs/research_reports/RR4402.html

Roehrig, G. H., & Kruse, R. A. (2005). The role of teachers' beliefs and knowledge in the adoption of a Reform-Based curriculum. School Science and Mathematics, 105(8), 412-422

Smith, M. S., & O'Day, J. (1990). Systemic school reform. Journal of Education Policy, 5(5), 233–267.

Summers, R. (2024). Appraising instructional materials from TeachersPayTeachers for features of NGSS design and nature of science representations. Research in Science Education, 54(3), 523-546.

Teaching Lab. (2024). On teacher mindsets and expectations. Teaching Lab.

Thomas-Browne, C. G., Boston, M. D., & Parke, C. S. (2023). Developing a tool to capture productive and unproductive mindsets about teaching mathematics to African American students: An initial review. Urban Education, 58(8), 1799-1826.

TNTP. (2018). The opportunity myth: What students can show us about how school is letting them down—and how to fix it. The New Teacher Project. https://tntp.org/tntp_the-opportunity-myth_web/

TNTP. (n.d.-a). Literacy – Reading Foundational Skills Observation Protocol. https://tntp.org/wp-content/uploads/Tools/observation-protocol-rfs.pdf

TNTP. (n.d.-b). Literacy – Reading Comprehension Observation Protocol. https://tntp.org/wp-content/uploads/Tools/classroom-observation-protocol-literacy.pdf

Appendix A: Survey Items

Baseline and Post-Survey Questions

Perception of Standards

Response options: Likert: Strongly disagree to Strongly agree (Strongly disagree/somewhat disagree/neither agree nor disagree/somewhat agree/strongly agree)

1. All of my students can master grade-level standards by the end of the year.
2. The standards are appropriate for the students in this class.
3. One year is enough time for students in this class to master the standards.
4. It's fair to expect students in this class to master these standards by the end of the year.

Expectations of Students

Response options: Likert: Strongly disagree to Strongly agree (Strongly disagree/somewhat disagree/neither agree nor disagree/somewhat agree/strongly agree)

1. Students of all ethnic or cultural backgrounds can be successful in my classroom
2. Students of all ethnic or cultural backgrounds are capable of solving problems by using critical thinking in my classroom
3. Students of all ethnic or cultural backgrounds are able to meet the expectations for higher-order skills in my classroom
4. Teachers should provide all students the opportunity to work with grade-level texts and tasks.

Student Need for Remediation

Response options: Likert: Strongly disagree to Strongly agree (Strongly disagree/somewhat disagree/neither agree nor disagree/somewhat agree/strongly agree)

1. Before students are asked to engage in complex learning tasks, they need to have a solid grasp of basic skills.
2. I try to keep in mind the limits of my students' ability and give them assignments that I know they can do so that they do not become discouraged.
3. It is not fair to ask students who are struggling with English to take on challenging academic assignments.
4. Grouping students of different levels of achievement for instruction may benefit some students, but it can undermine the progress that could otherwise be made by higher-achieving students.
5. Students who come into my classroom behind grade level will have a hard time succeeding.

Positive Regard for Curriculum

Response options: Likert: Strongly disagree to Strongly agree (Strongly disagree/somewhat disagree/neither agree nor disagree/somewhat agree/strongly agree)

1. The [subject] curriculum materials adopted by my school or district are well-suited to the needs of my students.
2. The [subject] curriculum materials adopted by my school or district offer students high-quality opportunities to learn.

3. The [subject] curriculum materials adopted by my school or district are well-organized and easy to use.
4. I like the [subject] curriculum materials adopted by my school or district.
5. The [subject] curriculum materials adopted by my school or district will help my students learn.

Leadership Support for Curriculum

Response options: Likert: Not at all to All the time (Not at all/Seldom/Sometimes/Often/All the time)

1. My school leaders attend the professional development [related to curriculum materials] with us.
2. My school leaders have created a shared vision for instruction that my [curriculum-related] professional development experiences is helping our school to implement.
3. My school leaders make sure I have access to all the materials and resources I need to implement our adopted curriculum.
4. My school leaders make time for my [curriculum-related] professional development.

Curriculum Barriers

Response options: Likert: Not at all to All the time (Not at all/Seldom/Sometimes/Often/All the time)

1. Instructional guidance in my school conflicts with the approach taken in [curriculum-aligned] professional development.
2. I sometimes feel pressure to teach in ways not aligned with the approach to instruction taken in [curriculum-aligned] professional development.
3. Sometimes I feel like my school/district puts up barriers to implementing the things I learn in [curriculum-related] PL.

Curriculum Use

Response options: Never use, sometimes (once a month), use often (once or twice weekly), use every day

1. Please indicate the extent to which you use [subject matter] curriculum materials adopted by your district.

Post-Survey Only Questions

Curriculum Mindset

Response options: Likert: Not at all to Extremely (Not at all/A little/Moderately/Very much/Extremely)

1. To what extent do you feel comfortable with students' uptake of the curriculum?
2. To what extent do you feel comfortable with sticking to the curriculum this semester?
3. To what extent do you feel your values and goals align with your use of the curriculum?
4. To what extent does the curriculum allow you to maintain meaningful relationships with all of your students?

Appendix B: Attrition Analysis

Table B1. Attrition in the sample, by teacher characteristics

	Randomized	*BaselineSurvey*	*Post-Survey*	*Observation*	*Student Data*
Grade Level					
PK-2	43	43	35	33	16
3-5	27	27	21	17	20
6-8	30	30	22	12	14
9-12	41	41	33	17	21
Missing	14	14	8	1	0
Subject					
ELA	80	80	69	58	46
Math	34	34	22	15	25
Other	41	41	28	7	1 (science)
SPED Teacher					
Yes	28	28	23	9	9
No	127	127	96	71	62
Total N	155	155	119	80	72

Differential attrition

To assess whether a teacher's treatment status is predictive of whether any outcome data was collected or missing, we created a binary indicator for the missing outcome and regressed this on the teacher's treatment status using a logistic regression. The coefficient for 'Treatment Indicator' represents the log-odds that the outcome data is missing based on the teacher's treatment status. For example, a coefficient of -1.040 means that the log-odds of post-survey data being missing decreases by 1.040 units when the teacher is in the treatment group compared to the control group. The standard errors for the post-survey treatment indicator indicates that the relationship between the teacher's treatment status and the likelihood of missing the post-survey is statistically significant ($p < 0.05$).

Table B2. Effects of Treatment on Attrition

Outcome Variables	Post-Survey (1)	Observations (2)	Student Outcomes (3)
Treatment Indicator	-1.040**	-0.260	-0.284
	(0.406)	(0.321)	(0.323)
Constant	-0.751	0.000	0.310

	(0.243)	(0.226)	(0.229)
Observations	155	155	155
Pseudo R^2	0.042	0.000	0.004

Note: *p<0.1; **p<0.05; ***p<0.01.

Appendix C: Observation Results by Instrument

Table C1. Effects of Treatment on Teacher Observation Scores – IPRT–M

	Teacher EOY Observation Scores: IPRT–M Instrument					
Variables	(1) Average of all IPRT-M indicators	(2) Depth	(3) Methods	(4) Prompting Thinking	(5) Engage	(6) Justification
Treatment Impact	0.189	1.029	-0.124	-0.560	0.285	0.317
	(.735)	(1.585)	(1.305)	(.610)	(2.641)	(0.146)
Constant	1.456	-2.309	0.703	7.549	3.026	-1.687
	(3.967)	(8.549)	(7.040)	(3.290)	(14.245)	(.790)
Observations	14	14	14	14	14	14
R^2	0.8887	0.9131	0.9073	0.9602	0.8662	0.9963
Adjusted R^2	-0.4469	-0.1291	-0.2049	0.4822	-0.7395	0.9519

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors are robust to heteroskedasticity. All models include controls for baseline survey results, when available and applicable, and fixed effects for district, grade-level, and special education teaching status.

Table C2. Effects of Treatment on Teacher Observation Scores – TNTP Literacy – Reading Comprehension Observation Protocol

	Teacher EOY Observation Scores: TNTP Literacy – Reading Comprehension Observation Protocol				
Variables	(1) Average of all indicators	(2) Culture of Learning	(3) High Quality Texts	(4) High Quality Questions & Tasks	(5) Student Ownership
Treatment Impact	0.051	0.330	-0.008	-0.053	-0.066
	(.253)	(.305)	(.306)	(.322)	(.300)
Constant	1.906	3.161*	2.111	1.194	1.157
	(1.195)	(1.298)	(1.228)	(1.553)	(1.345)
Observations	46	46	46	46	46
R^2	0.6464	0.6147	0.6871	0.4840	0.5955

Adjusted R^2	0.3880	0.3331	0.4584	0.1069	0.2999

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors are robust to heteroskedasticity. All models include controls for baseline survey results, when available and applicable, and fixed effects for district, grade-level, and special education teaching status.

Table C3. Effects of Treatment on Teacher Observation Scores – TNTP Literacy – Reading Foundational Skills Observation Protocol

	Teacher EOY Observation Scores: TNTP Literacy - Reading Foundational Skills Observation Protocol						
Variables	(1) Summary Rating	(2) Average of all indicators	(3) Culture of Learning	(4) Aligned Content	(5) Teacher-Directed Instruction	(6) Student Practice	(7) Assessment & Differentiation
Treatment Impact	0.208	0.268	0.488	0.292	0.493	0.446	-0.381
	(.632)	(.398)	(.513)	(.571)	(.625)	(0.319)	(.387)
Constant	10.799	9.217	8.512	5.027	11.597	10.891*	10.056
	(6.039)	(3.979)	(3.780)	(4.944)	(5.931)	(3.407)	(4.088)
Observations	12	12	12	12	12	12	12
R^2	0.3865	0.4832	0.6981	0.3690	0.3734	0.6742	0.5996
Adjusted R^2	-0.3497	-0.1370	0.3358	-0.3882	-0.3786	0.2832	0.1191

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors are robust to heteroskedasticity. All models include controls for baseline survey results, when available and applicable, and fixed effects for district, grade-level, and special education teaching status.

Appendix D: Student Outcome Robustness Check

Table D. Effects of Treatment on Student Performance – Alternate Specifications

	Student EOY Scores			
Variables	(1) No SPED FE	(2) SPED Teachers Dropped	(3) Demographic Controls	(4) Multilevel Model
Treatment Impact	-0.113	-0.082	0.017	-0.032
	(0.151)	(0.149)	(0.136)	(0.110)
Constant	0.091	0.113	0.206	-0.034
	(0.082)	(0.079)	(0.139)	(0.119)
District FE	x	x	x	
Grade-Level FE	x	x	x	
Subject FE	x	x	x	
SPED FE				
Observations	2735	2652	526	2735
R^2	0.016	0.022	0.138	
Adjusted R^2	0.006	0.012	0.085	

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors are clustered at the teacher level.