TEACHING LAB PROFESSIONAL LEARNING SERIES:
SUPPORTING TEACHERS TO DELIVER HIGH-QUALITY, EQUITY-ORIENTED, AND CURRICULUM-ALIGNED MATHEMATICS INSTRUCTION

June 2021

Dr. David Blazar (External Evaluator)
with Frances Henderson, Doug McNamara, and Blake Turner (Members of Research Team)

Summary of Research Design, Findings, and Recommendations

This report analyzes how Teaching Lab's Professional Learning Series was delivered to mathematics educators across New Mexico in spring 2021, as well as associated changes in teachers' knowledge of math pedagogy, self-efficacy to deliver high-quality instruction, and teacher/administrator-reported and outside observers' assessments regarding the quality of teachers' instruction. As outside evaluators, we designed a mixed-methods research study in which we leveraged quantitative survey data from participating teachers and school administrators/principals, observations of mathematics lessons, and focus group discussions. Analyses of the quantitative survey data allow us to examine pre/post changes in desired outcomes, while qualitative analyses of classroom observations and focus group discussions help us to identify potential mechanisms underlying changes in outcomes, provide targeted recommendations for improving future delivery of the professional learning series content, and inform continued research.

In January and February 2021, Teaching Lab coordinated with staff at the Math and Science Bureau within the New Mexico Public Education Department (PED) to recruit schools and teachers to participate in a set of professional learning activities aimed at supporting the delivery of high-quality mathematics instruction. Activity 1 focused on developing pedagogical knowledge and instructional strategies for equitable mathematics practices. Content delivered synchronously and asynchronously across eight sessions was aligned to the New Mexico Common Core State Standards for Mathematics, the Standards for Mathematical Practice, the National Standards for Quality Online Learning, the New Mexico Instructional Scope, and the PED-adopted balanced system of assessments. The content also aimed to support teachers in their delivery of curricula-aligned instruction, with elementary teachers and school teams using *EngageNY* and secondary teachers and school teams using *Illustrative Math*. Activity 2 was designed to leverage the New Mexico Instructional Scope to support acceleration and re-entry for K-12 instruction post Covid-19, as well as just-in-time instructional support for *all* students to be able to access grade-level content. Activity 2 occurred over three sessions. In total, 48 math teachers across 13 schools, 10 school leaders, and 14 other participants (e.g., non-math teachers, instructional support staff, parent liaisons) participated in one or both activities; roughly half of the teachers participated in both activities.

In summary, we found:

- **Attendance was strong**, with an average of 75% of sessions attended among those individuals who decided to take part in the professional learning series. However, due to the nature of the Covid-19 pandemic and the return to school in mid-spring, roughly half of the teachers and school teams that initially expressed interest did not actually participate at all. Because cross-school professional learning series teams were created before this time, there was some disruption to learning due to the initial lack of take-up.
- Math teachers who participated showed substantively **meaningful changes in their**

knowledge of math pedagogy, self-efficacy at delivering high-quality instruction, and self-reported practices related to high expectations for student learning. Standardized differences in pre- versus post-implementation scores generally were around 0.4 standard deviations (SD), with even larger standardized differences for teachers' knowledge of math pedagogy (as large as 1 SD). This latter difference is equivalent to moving the median teacher from the 50th to the 84th percentile in the distribution of effectiveness. As a further point of comparison, the difference in teacher- and administrator-reported metrics of effectiveness between teachers in their first four years in the class versus veteran teachers were around 0.5 SD. As such, it is reasonable to infer that standardized differences associated with participation in the professional learning series are on par with or larger than the difference between being a novice versus a veteran teacher. **Positive change scores were quite similar across the two professional learning activities, but larger among secondary teachers relative to elementary teachers**. Pre/post implementation changes in principal reports of teachers' practice and overall instructional effectiveness also were positive and, in some cases, substantively significant; however, these differences are not statistically significantly different from 0.

- Two teachers who volunteered to record their math lessons also exhibited **concrete differences in the quality and rigor of their instruction**. Compared to lessons captured prior to the start of the professional learning series, lessons captured at the end of the program tended to include more emphasis on student questioning, as well as stronger use of manipulatives to support students' engagement in rigorous, conceptually demanding, and grade-level tasks.
- Five teams that participated in focus group discussions at the end of training described **strong interest in the content and materials**, particularly around equity-oriented lenses about how to engage *all* students in the work, and scaffolds for accelerating learning. In these discussions, teachers and school leaders also described a strong **appreciation and interest in working in cross-school teams** to leverage expertise across the state.

Based on these findings and patterns, we make the following recommendations:

- **Benefit of continued work with Teaching Lab:** Our findings suggest that teachers and school teams engaged meaningfully in the professional learning series—including both Activities 1 and 2—and that this engagement was associated with positive changes in desired outcomes. From these data, we infer that additional schools and teachers in New Mexico could benefit from these experiences.
- **Areas of Improvement in Content:** Data captured primarily from the focus group discussions also reveal ways in which the content and materials driving the professional learning series could be improved. We expand on this discussion at the end of the report and briefly list potential improvement areas here: (1) lengthen Activity 2 to allow for deeper engagement; (2) allow for longer planning and recruitment period to gauge interest and gain momentum for full participation; (3) as facilitators continue to build knowledge of the Teaching Lab content, move away from some of the scripts and allow participants to drive discussions; and (4) balance equity-oriented content related to *access* to rigorous mathematics, which often was how participants defined "equity", with stronger emphasis on *cultural responsiveness* that also is part of Teaching Lab's equity-oriented mission but not often mentioned by teachers or school leaders.
- **Areas for Continued Research:** The current study identified positive trends in desired

outcomes, while also raising additional questions that may be areas for ongoing inquiry. We expand on this discussion at the end of the report and briefly list potential areas for continued research here: (1) probe factors that may stand in the way of high attendance and engagement turning into changes in desired outcomes, particularly for elementary teachers where attendance and interest was quite strong but where changes in outcomes were not as large as they were for secondary teachers; (2) increase sample size to allow for deeper subgroup analyses, including school-level trends, school level-by-activity trends, and activity-by-curriculum trends; (3) examine whether dosage/number of sessions makes a difference in changes in desired outcomes; (4) collect outcomes for a comparison group that did not engage in Teaching Lab activities—and ideally randomly assign individuals to the treatment versus control group—in order to ensure that positive pre/post changes in outcomes are *caused* by the professional learning series; and (5) collect observations of instructional practice for all teachers, as well as student outcomes (e.g., test scores, surveys) in order to examine impacts of programming on students' classroom experiences and academic development.

Introduction

This report analyzes how Teaching Lab's Professional Learning Series was delivered to mathematics educators across New Mexico in spring 2021, as well as associated changes in teachers' knowledge of math pedagogy, self-efficacy to deliver high-quality instruction, and teacher/administrator-reported and outside observers' assessments regarding the quality of teachers' instruction. As outside evaluators, we designed a mixed-methods research study in which we leveraged quantitative survey data from participating teachers and school administrators/principals, observations of mathematics lessons, and focus group discussions. Analyses of the quantitative survey data allow us to examine pre/post changes in desired outcomes, while qualitative analyses of classroom observations and focus group discussions help us to identify potential mechanisms underlying changes in outcomes, provide targeted recommendations for improving future delivery of the professional learning series content, and inform continued research.

The professional learning series and associated research is grounded in and driven by a theory of change, where (1) changes in knowledge of math pedagogy paired with (2) changes in teachers' mindsets—namely their self-efficacy at delivering high-quality instruction—will (3) turn into specific changes in instructional practice (e.g., delivery of rigorous grade-level content, holding students to high expectations for learning), and (4) ultimately into changes in students' classroom experiences and academic performance. In this study, we measure changes in knowledge of pedagogy, self-efficacy, and practice associated with participation in the professional learning series. While we do not measure student outcomes directly in this study, we infer that changes in proximal outcomes at the teacher level likely translate into changes in more distal outcomes capturing students' experiences and outcomes.

Teaching Lab Professional Learning Series

Teaching Lab is a non-profit organization devoted to teacher professional development, particularly in the realms of curriculum-aligned instruction and educational equity. In its creation of professional learning content, Teaching Lab attends to the "core features" of effective programming from scholarly research, which include: (1) intensive and sustained durations; (2) focus on discrete skills; and (3) application of these skills in context/in teachers' own classrooms (Darling-Hammond et al., 2009; Desimone et al., 2002; Garet et al., 2002). In other words, the scholarly literature describes effective professional learning for teachers as activities that approach the work as a dynamic, active process where teachers may engage directly with student work, obtain direct feedback on their instruction, or review materials from their own classrooms.

In early 2021, the New Mexico Public Education Department (PED) partnered with Teaching Lab to design and facilitate a professional learning series that was delivered to individuals in 13 schools and nine school districts across the state. Teaching Lab designed two learning activities which focused on improving teachers' use of equitable mathematics practices in their classrooms and supporting students to engage in rigorous, grade-level tasks.[1] Participants had the option to attend Activity 1, Activity 2, or both activities. These professional learning activities were designed primarily for

[1] Teaching Lab also designed and delivered a leadership series targeted at school leaders only. This series provided wrap-around support for leaders through two synchronous sessions at the beginning and end of the activities, along with asynchronous opportunities aligned to each session within Activity 1 and/or 2. We did not collect outcome measures for participating leaders, as the primary goal of this supplemental leadership series was to improve leaders' support for teachers to deliver high-quality instruction. Several measures related to teachers' ability to deliver high-quality instruction *are* included in this report; we infer that changes in these measures may be attributable to the combined teacher-facing activities *and* to the leadership series.

supporting math teachers (N = 48 in our final sample, with 29 math teachers in Activity 1, 39 math teachers in Activity 2; 20 math teachers participated in both activities). School leaders (N = 10) also participated as a means of supporting and facilitating growth for their teachers, and building a school climate conducive to ongoing instructional improvement. With this same goal in mind, in some instances, teachers in subject areas other than math and other members of the school community (e.g., parent liaisons) also participated (N = 14). In total, 72 people participated in the learning series (see Table 1).[2]

Activities 1 and 2

Activity 1 was designed to develop teachers' pedagogical content knowledge and use of rigorous instructional strategies. Teaching Lab aligned Activity 1's content to the New Mexico Common Core State Standards for Mathematics, the Standards for Mathematical Practice, the National Standards for Quality Online Learning, the New Mexico Instructional Scope, and the New Mexico PED's balanced system of assessments. Activity 1 comprised a mix of eight synchronous and asynchronous sessions. As shown in Table 1 below, 34 participants (29 teachers) from four schools took part in Activity 1. Teaching Lab designed Activity 1 to support multiple modalities of instruction (i.e. virtual, hybrid, in-person). Teaching Lab attended to and relied upon two proven curricular packages—*EngageNY* for elementary teachers and *Illustrative Mathematics* for secondary teachers—into the training. Topics in the sessions included:

- Attending to rigor and incorporating the eight effective mathematics teaching practices developed by the National Council of Teachers of Mathematics;
- Strengthening students' conceptual understanding with rich activities;
- Affirming mathematics learners' identities;
- Challenging spaces of marginality by honoring student voice and centering student knowledge and expertise;
- Introducing and incorporating Universal Design for Learning principles into planning and instruction; and
- Developing learning goals which support subsequent instructional decisions.

Leveraging New Mexico's Instructional Scope, Activity 2 supported acceleration and re-entry for K-12 students, particularly students who may have fallen behind because of the Covid-19 pandemic. Acceleration involves teachers identifying the most important grade-level standards students need to understand and then strategically prioritizing that learning to ensure all students have access to grade-level instruction. Activity 2 comprised three sessions, which included topics such as:

- Developing instructional routines;
- Providing just-in-time supports;
- Understanding and implementing high-leverage instructional moves; and
- Meaningfully using practice problems.

---

[2] At the outset of the project, Teaching Lab and PED initially set out to recruit 50 K-12 school teams, which included school instructional leaders and administrators, as well as at least 80% of a school's staff responsible for mathematics instruction. Following recruitment, 131 individuals expressed interest and were listed on the professional learning series rosters. However, several schools and several teachers within participating schools dropped out of the program before/at the very start of training, due to competing commitments and constraints, driven in particular by Covid-19 and the return to in-person learning.

Participant Attendance

Table 1 summarizes the attendance of math teachers, administrators, and other participants in both activities. In total, 72 participants attended Activity 1, Activity 2, or both activities. Of those, 48 (67%) were math teachers, who are the primary focus of our analyses of pre/post changes in outcomes. Attendance was similar between math teachers and other participants, and similar across Activities 1 and 2 for these individuals (roughly 70% or more of sessions attended). Participation was slightly lower for school administrators, who attended roughly half of Activity 1 sessions (out of 8 total) and roughly two-thirds of Activity 2 sessions (out of 3 total). Though not shown in table 1, attendance was almost identical between elementary and secondary teams.

Table 1. Participation in the Teaching Lab Professional Learning Series

	Math Teachers	School Administrators	Other Participants
All (13 Schools)			
N	48	10	14
Average Number of Sessions Attended (up to 11)	5.2	2.5	3.3
% of Total Sessions Attended	72%	65%	86%
% Attended at least Half of Sessions	73%	80%	79%
Activity 1 (4 Schools)			
N	29	2	3
Average Number of Sessions Attended (out of 8)	5.5	4.5	5.3
% of Total Sessions Attended	69%	56%	67%
% Attended at least Half of Sessions (i.e., 4+)	88%	90%	93%
Activity 2 (12 Schools)			
N	39	8	11
Average Number of Sessions Attended (out of 3)	2.4	2	2.7
% of Total Sessions Attended	79%	67%	91%
% Attended at least Half of Sessions (i.e., 2+)	88%	90%	93%

Note: "Other Participants" include district resource officers and curriculum specialists, coaches, teachers in subject areas other than math, and parent liaisons.

Characteristics of Teachers in Analysis Sample

In Table 2, we describe demographic characteristics of the math teachers (N = 48) who contribute to our primary analyses examining pre/post changes in key outcomes. (See below for a description of these measures.) We captured demographic characteristics from the same survey in which we captured teacher-reported mindsets, practices, and pedagogical knowledge; we did not send this survey to school leaders and non-math teachers, and thus do not have demographic characteristics for them. Our rationale for this decision was that school leaders helped *provide* some of the outcome data on teachers and facilitate teacher learning, but who were not the primary target for improving mindsets, practices, or knowledge. Similarly, we excluded the set of non-math teachers, as the outcome measures ask specifically about math knowledge and math instruction.

Across both activities, we see similar demographic breakdowns. The majority of participants identified as female (roughly 80%) and Hispanic (roughly 55%), overall and in each activity. Slightly more secondary teachers participated than elementary teachers (52% versus 48%). However, compared to secondary teachers, a much larger percentage of elementary teachers participated in both activities (41% of elementary teachers versus 16% of secondary teachers); this is why there is a greater share of elementary teachers in Activities 1 and 2, while elementary teachers make up a slightly smaller share of the overall sample. Across both activities, participants had similar average years of math teaching experience (between 9 and 10 years). One-third of the full sample were novices in their first through fourth years, while two-thirds were veterans; there was a slightly higher percentage of novice teachers in Activity 2. Overall, the majority of participants had three or fewer years of experience with their current curricula; only seven of the 48 participants (15%) had five or more years of experience with their current curricula.

Table 2. Characteristics of Participating Math Teachers in Quantitative Analyses

	All	Activity 1	Activity 2
Demographics			
Female (%)	79%	72%	87%
Hispanic (%)	54%	55%	51%
White (%)	42%	48%	44%
Race Missing/Other (%)	8%	3%	10%
Elementary (%)	48%	52%	56%
Secondary (%)	52%	48%	44%
Math Teaching Experience (Years)	9.8	8.6	9.6
Novice Teacher (i.e., 1 to 4 years; %)	33%	34%	38%
Veteran Teacher (i.e., 5 plus years; %)	67%	66%	62%
Experience with Current Curriculum (Years)	2.4	2.5	2.4
Has Outcome Measures			
At Least 1 Teacher Survey (%)	100%	100%	100%
Both Pre/Post Teacher Surveys (%)	67%	62%	72%
At Least 1 Administrator Survey (%)	94%	97%	92%
Both Pre/Post Administrator Surveys (%)	50%	55%	51%
N	48	29	39

Note: Activity 1 and Activity 2 samples are not mutually exclusive. Teacher characteristics are only collected from math teachers, which is the sample of participants that were sent and asked to complete a teacher survey about their work as math teachers and knowledge of math pedagogy.

Research Questions, Data Sources, and Analyses

In this study, we ask two research questions:

1. *To what extent did teachers who participated in Teaching Lab's training improve in their knowledge of math pedagogy, self-efficacy to deliver high-quality instruction, and teacher/administrator-reported and outside observers' assessments regarding the quality of instruction?*
2. *How was the program implemented and how might it be improved in the future?*

In order to answer these two research questions, we collected and analyzed data from four sources: (1) pre- and post-implementation surveys of teachers capturing information on their knowledge of math pedagogy, self-efficacy to deliver high-quality instruction, and self-reported practices related to holding students to high expectations for learning; (2) pre- and post-implementation surveys of administrators/school leaders rating individual teachers on their efficacy at delivering high-quality instruction; (3) pre- and post-implementation video observations from two teachers; and (4) focus group discussions with five teams. The mix of quantitative data (first two data sources) and qualitative data (latter two data sources) allow for a mixed-methods exploration of *how* the professional learning series was implemented, *whether or not* training was associated with changes in teacher outcomes, and *potential mechanisms* that might explain these changes (Creswell & Clark, 2017). We describe each source and relevant analyses below.[3]

Quantitative Data and Analyses to Capture Changes in Teacher Outcomes

Prior to the start of training and then again at the end of training, we collaborated with Teaching Lab to administer surveys to each participating teacher and to a school leader with knowledge of that teacher's classroom work. For the teacher and administrator surveys, items were the same pre- versus post-implementation, allowing us to examine changes in outcomes over time.

Teacher Survey. For the teacher survey, questions included background demographic information (i.e., see Table 2) and items aimed at capturing three distinct dimensions of high-quality teachers and high-quality teaching:

1. Teachers' **knowledge of math pedagogy** (26 multiple-choice items; internal consistency reliability [Cronbach's alpha] = 0.80);
2. Teachers' **self-efficacy** at delivering various dimensions of high-quality instruction (6 items on 1 to 5 Likert scale; alpha = 0.87);
3. The extent to which teachers build strong working relationships with students and families, and ultimately hold students to **high expectations** in their instruction (11 items on 1 to 5 Likert scale; alpha = 0.71).[4]

---

[3] To participate in the research component of this project, participating teachers and school administrators/principals signed several consent forms. All participants signed consents with Teaching Lab acknowledging and agreeing to their participation in the professional learning series, and completion of surveys. Teachers who submitted videotaped lessons also signed consent forms with Teaching Lab approving use of those videos for research purposes. Finally, teachers and school administrators/principals who participated in the focus group discussions signed consent forms administered by the research team and approved by the Institutional Review Board at the University of Maryland, College Park.

[4] We originally envisioned items within the High Expectations domain as two distinct constructs: relationships with students and families versus high expectations. However, analyses suggest that they perform better as a single construct, which is consistent with the idea that strong working relationships and high expectations are key features of culturally responsive teaching (Gay, 2002)

Teaching Lab designed the math pedagogy assessment prior to the start of this study, though we adapted it to include a parsimonious set of items that were most closely aligned to the professional learning series prepared for New Mexico. The research team adapted the other teacher survey items and constructs from large-scale research projects, including the National Center for Teacher Effectiveness (e.g. Hill et al., 2015), focused on capturing multiple dimensions of teacher and teaching effectiveness. We focused on items that had strong measurement properties and that were closely aligned to Teaching Lab's work aimed at increasing both teachers' mindsets and practices related to the delivery of rigorous and equity-oriented mathematics content. (See Appendix 1 for the full teacher survey; we exclude items capturing teachers' knowledge of math pedagogy, as these are proprietary to Teaching Lab.)

In the current study and sample, we analyzed the measurement properties of each construct, and found them to be strong. Internal consistency reliability measures (i.e. Cronbach's alpha) are at or above traditional thresholds in education research (Taber, 2018) and are consistent with reliability estimates of the same constructs in other samples (Hill et al., 2015). We created composite measures for each of the dimensions by reverse coding some items so that all items had the same valence, and then averaging across items.

We administered the same survey prior to the start of training and then again at the end of training. One hundred percent of math teachers completed the survey at least once (pre- or post-implementation), and roughly two-thirds of math teachers completed the survey at both time points (see Table 2).

Administrator/Principal Survey. The research team also developed and administered a survey to school leaders aimed at capturing an outside perspective on the quality of teachers' work and instructional practice. Items were adapted from surveys of school administrators in projects of teacher quality generally (Harris & Sass, 2014; Jacob & Lefgren, 2008) and changes in teacher practice in the context of professional development (Blazar & Kraft, 2015; 2019).

This survey asked administrators to rate each participating teacher on their overall quality of instruction, as well as their effectiveness at building student conceptual understanding, delivering math content to all learners, and connecting with their students' lives and cultures out of the classroom (4 items on 1 to 7 Likert scale; alpha = 0.98). An alpha of 0.98 indicates that the four principal ratings were highly correlated and capture the same underlying construct (i.e., teachers' overall quality of instruction). Thus, to simplify analyses, we created a composite principal rating that averaged across the four individual ratings. (See Appendix 1 for the full administrator/principal survey.)

As with the teacher survey, we asked school administrators to rate their teachers both prior to the start of training and then again at the end of training. The vast majority of math teachers (over 90%) had at least one administrator rating, and roughly half of math teachers had administrator ratings at both time points (see Table 2).

Quantitative Analysis. Our primary interest in collecting and analyzing the teacher and administrator/principal surveys was to understand how teachers' knowledge of math pedagogy, self-efficacy at delivering high-quality instruction, and teacher/administrator-reported assessments regarding the quality of teachers' instruction changed over the course of their participation in the Teaching Lab training.

To estimate these pre/post changes in outcomes, we used a regression framework to test for statistically significant differences in means. Regression analyses (as opposed to more simple t-tests) allowed us to include school fixed effects as a way of accounting for how school characteristics might affect all teachers *within* that school and that differentiate teachers *between* schools. For instance, teachers within the same school may receive common guidance for instruction, planning, etc. from their school administration; however, this guidance may differ between schools. Similarly, administrators from different schools may have different standards and criteria for rating teachers; however within a school, a single administrator would apply the same standards and criteria when rating their teachers. These are examples of "reference bias" (Duckworth & Yeager, 2015) that can be accounted for by conditioning on teachers' schools (West et al., 2016).

To conduct these analyses, we first created a dataset in which we stacked pre- and post-implementation survey data, such that each teacher had up to two rows in the final dataset. Then, we predicted each of the teacher- or administrator/principal-reported constructs as a function of a dummy indicator for the post time period. As noted above, we also included school fixed effects. To account for multiple observations per teacher, we clustered standard errors at the teacher level. Over half of teachers had both pre- and post-implementation surveys (see Table 2), though we did not restrict our analysis sample to those teachers with responses at both time points. Robustness tests that do make this restriction identify almost identical patterns to those shown below.

Qualitative Data and Analyses to Understand Program Implementation and Potential Mechanisms Underlying Quantitative Relationships

We supplemented our quantitative analyses with qualitative observation of classroom instruction and focus group discussions. Together, these additional data sources help illustrate, expand, and elaborate patterns that emerged from the quantitative results (Hill et al., 2008), and afford open-ended exploration of potential mechanisms (including both strengths and weaknesses) underlying changes in teacher outcomes, with the intent of developing themes from the data that were not visible in the quantitative results (Creswell & Clark, 2017).

Video Observations. Similar to the teacher and administrator/principal surveys, we collected videotaped lessons prior to the start of teachers' participation in the professional learning series and then again at the end in order to examine changes in desired outcomes—in this case, instructional practice. Collection of videotaped lessons aligns with a primary goal of Teaching Lab and of PED to engage in professional learning for the purpose of improving the quality of classroom instruction.

Due to the intensive nature of collecting and analyzing classroom observations (Hill & Grossman, 2013) as well as our goal in this portion of the study to illustrate patterns and expand on the quantitative results, we focused data collection on a limited number of teachers. With the help of Teaching Lab, we recruited teachers from the full sample by sending an email invitation describing the purpose of these data and offering a $100 Amazon gift card for participating. Three teachers agreed; however, with the return to in-person learning mid semester, one of these teachers was not able to record a second lesson. Therefore, we focus our analyses on the two teachers who contributed both pre- and post-implementation lessons. Both were elementary teachers who engaged in Activities 1 and 2. All four lessons (i.e., pre/post lessons for each of the two teachers) were recorded via Zoom with the main screen showing teachers' instruction.

Before observing these lessons, we created a video observation protocol aligned to the professional learning series content (see Appendix 2). The protocol includes two broad sections focused on mathematical practices and on equitable practices, each of which has subdomains and specific "look-fors." The section of mathematics practices includes subdomains aligned to the New Mexico Standards of Mathematical Practice, including a focus on mathematical modeling and use of mathematical tools to foster student sensemaking. Subdomains within equity practices are oriented around Teaching Lab's definition of equity, including increased access to grade-level mathematics, holding students to high expectations, and integration of discussion to unearth diverse viewpoints and to acknowledge students' unique racial, ethnic, and cultural backgrounds. Because of the open-ended nature of our qualitative observations of classrooms, we designed the protocol to have observers identify whether or not they saw each of these "looks fors" and to describe how they were instantiated in teachers' instruction. The protocol did not ask observers to score each "look for" on a numeric rating scale.

All four researchers observed all lessons, with each of three researchers assigned to focus on a subset of "look-fors." One additional researcher was assigned as the lead for each lesson, paying attention to all "look-fors," and responsible for leading group discussion of that lesson and for writing a memo of that discussion that identified themes within that specific lesson. Notes on the observation protocol and the memos are the data source that we ultimately analyzed to look for themes across lessons.

Focus Group Discussions. Finally, we conducted five focus group discussions that included participants from the same school or team. To recruit schools and teams for these discussions, we collaborated with Teaching Lab to identify those that had engaged with the full professional learning series, had high attendance, and also were interested in sharing their experiences in discussion format. Specifically, Teaching Lab emailed individual teachers to gauge interest, and then followed up with individual administrators/principals where teacher interest and attendance was strong. The research team then emailed these teams to confirm interest in participating, identify a day/time that worked best for the group, and share consent forms. Each school team that participated in a focus group discussion received a $200 Amazon gift card.

We developed a protocol to guide the focus group discussions, with topics including identifying specific changes in mindsets and beliefs as a result of participating in the professional learning series, elements of the content and facilitation that they found to be the most effective, and areas of the training that could be improved upon (see Appendix 3). We met with each team for approximately one hour on Zoom. Each participant was provided the opportunity to answer each of the protocol questions, but not all participants opted to answer each question. We also allowed the conversation to move beyond the immediate questions posed in the focus group discussions in instances where the group wanted to share related insights.

During the discussion, one researcher led the conversation while another took notes. We also recorded the conversations. Following each discussion, the notetaker revisited notes in order to write up a memo capturing themes that emerged and specific quotations from individual teachers or school leaders. Like with the video observations, these memos are the data source we analyzed to look for themes across focus group discussions.

Qualitative Analysis. For both the videos and focus group discussions, we took a deductive approach to search for and identify themes across the data that helped to illustrate and expand upon findings from the quantitative analyses (Creswell & Clark, 2017). Each member of the research team individually reviewed the videos, attended or viewed the videos of each group discussion, and then reviewed the written memos for each in an effort to identify common themes. For the videos of mathematics lessons, our primary goal was to identify potential changes in practices pre- versus post-implementation of the professional learning series. For the focus group discussions, we were primarily interested in identifying themes regarding participants' instantiation of learning in their own schools and classrooms, the overall strengths of the professional learning series, and potential weaknesses and areas for improvement.

Given our goal of using the qualitative data in an open-ended fashion, we took a multi-step process to thematic analysis. First, each researcher identified trends *within* a given video observation or focus group discussion. Second, when the full research team met to review a given lesson or focus group discussion, we identified trends *across* researchers. Third, after reviewing data collected from multiple lessons or focus groups, the full research team identified common *themes across lessons or focus group discussions*. Because we conducted this work collaboratively as a full research team, we did not check for rater agreement of themes.

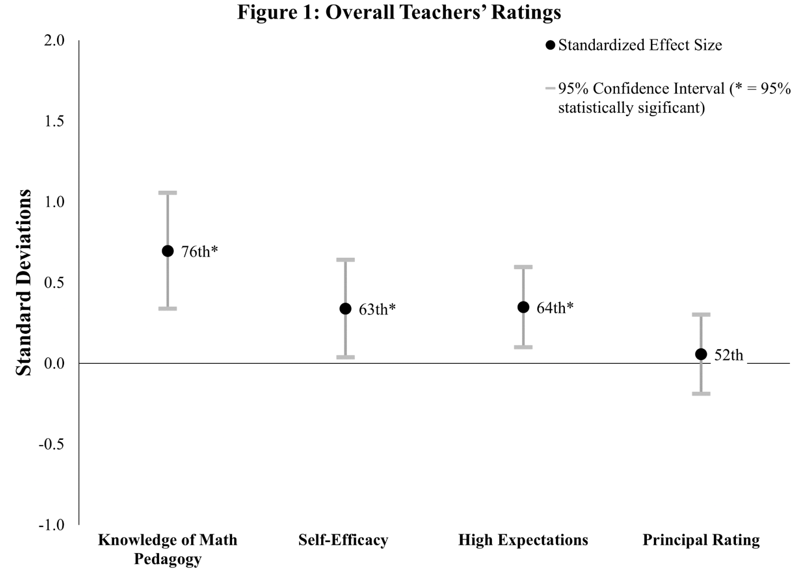
Findings

Surveys: Changes in Teacher Knowledge of Math Pedagogy, Self-Efficacy, and Practice

First, we present results from the quantitative analyses examining changes in teacher- and administrator/principal-reported survey constructs. We summarize patterns in figure form, starting by pooling data across activities and grade levels (Figure 1); we also compare results for Activity 1 versus Activity 2 (Figure 2), and elementary versus secondary teachers (Figure 3). The pooled results are appealing, as they include the largest possible sample of teachers, whereas subgroup analyses have smaller sample sizes that can lead to more limited statistical power to detect relationships that are statistically significantly different from 0. Further, while there are distinct differences between the two activities (e.g., number of sessions, instructional focus), close to half of the teachers who contribute to our quantitative analyses participated in both activities; thus, it is more difficult to tease out the unique role of one activity versus the other.

In each figure, we are interested in quantifying how much change we observe between the start and end of the professional learning series in teachers' *knowledge of math pedagogy, self-efficacy* to deliver high-quality instruction, and teacher-reported practices related to holding students to *high expectations*, and administrator/*principal ratings* that capture teachers' overall instructional effectiveness. To allow apples-to-apples comparisons among these four measures that were captured on different scales, we report all results as standardized differences. To do so, we divide the change in a given measure by its standard deviation in the pre-implementation period (see Y axis); an estimate of 0 indicates no change. To aid in interpretation, we further interpret the standardized changes on a percentile scale, reflecting how far up the distribution the median teacher moves from the start to end of training. We report these more readily interpretable results to the right of each black dot/point estimate. As a further point of comparison, we note that the difference in teacher- and administrator-reported metrics of effectiveness between teachers in their first four years in the class versus veteran teachers were around 0.5 SD. Finally, each figure adds vertical bars that reflect 95% confidence intervals; when these intervals do not cross 0, we conclude that the change is statistically significantly different from 0.

Figure 1: Overall Teachers' Ratings



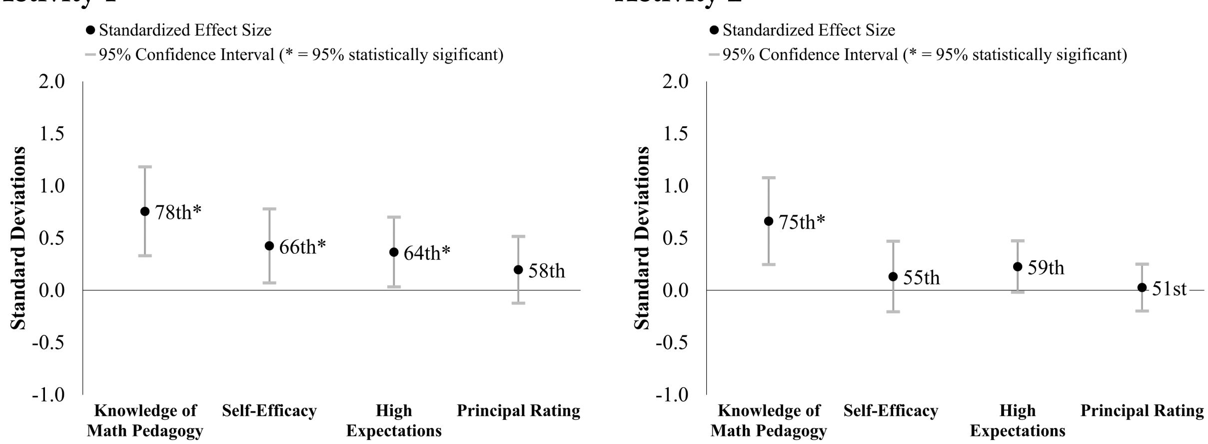
Overall Trends. When pooling across all teachers (see Figure 1), we observe positive standardized differences across all four teacher outcome measures. We find statistically significant changes in measures of teachers' *knowledge of math pedagogy, self-efficacy* to deliver high-quality instruction, and self-reported practices that hold students to *high expectations*. We observe the largest change in the first measure, of 0.7 SD; in other words, on average, the median teacher moved to roughly the 76th percentile in the distribution of their knowledge of math pedagogy. This estimate also is larger than the difference between novice and veteran teachers in our sample (roughly 0.5 SD). Estimates for teacher-reported *self-efficacy* and *high expectations* are very similar to each other, with a standardized change of roughly 0.35 SD; this estimate is equivalent to an average teacher moving from the 50th to the 63rd/64th percentile in the distribution of effectiveness. While positive, the change in administrator/principal reports of changes in the quality of teachers' instruction is smaller and not statistically significantly different from 0.

Trends by Activity. Next, we examine and report pre/post changes for teachers who participated in Activity 1 versus Activity 2 (see Figure 2). Of the 48 math teachers who contributed to these analyses, 20 participated in both activities, while nine participated just in Activity 1 and 19 participated just in Activity 2. Therefore, we caution readers from interpreting these results as the unique role of each activity on its own. It may be that patterns reflect the cumulative role of the two activities together.

Despite substantive differences in the two activities (e.g., instructional focus, number of sessions), we find patterns of results that are quite similar across the two, and quite similar to pooled differences shown in Figure 1. Specifically, we find positive and substantively meaningful changes in the three constructs from the teacher survey. For *knowledge of math pedagogy*, the standardized change is quite close across the two activities. For *self-efficacy* and *high expectations*, changes are slightly larger for Activity 1 than for Activity 2, and only statistically significantly different from 0 for the Activity 1 subsample. For the principal survey, change scores in their ratings of teachers' overall effectiveness are positive for both activities but not statistically significantly different from 0, as we also found in the pooled results. For Activity 1, the standardized difference in the principal survey rating of 0.20 SD is substantively meaningful, as it reflects a move from the 50th to the 58th percentile in the distribution of effectiveness.

Figure 2: Teacher Ratings by Activity

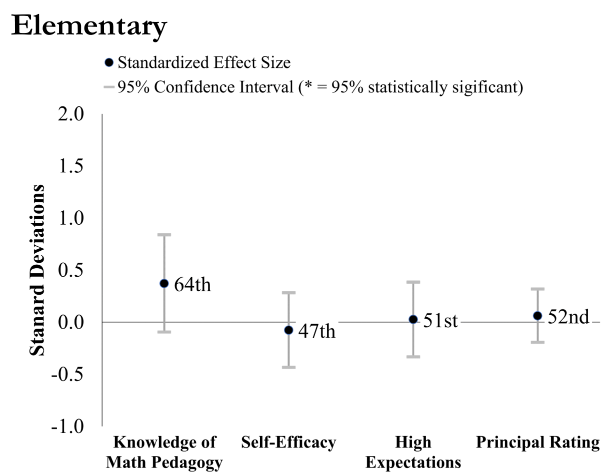
Activity 1 **Activity 2**

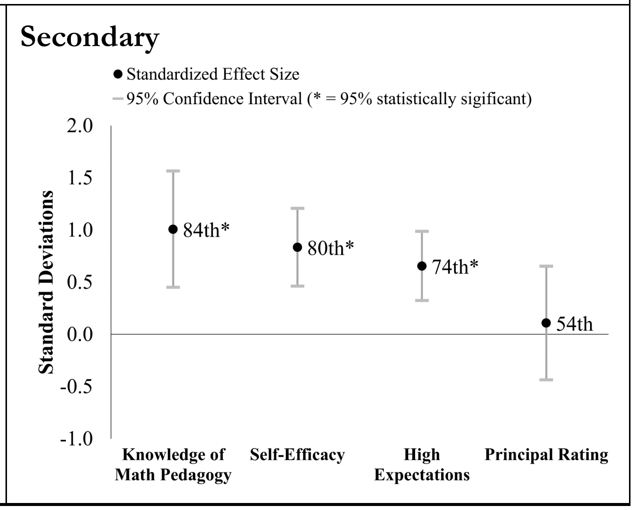


Trends by School Level. In Figure 3, we report changes by school level (i.e., elementary versus secondary). We do not further disaggregate by activity, as sample sizes would be too small to warrant substantive interpretations. Here, we observe substantively meaningful differences across grade levels. For secondary teachers, the standardized changes in all of the teacher survey measures are larger than in the pooled sample. For example, we observe a standardized difference of 1 SD for *knowledge of math pedagogy*, 0.83 SD for *self-efficacy*, and 0.65 SD for *high expectations*. The first estimate is equivalent to moving the median teacher to the 84th percentile in the distribution of knowledge, and the last estimate is equivalent to moving the median teacher to the 74th percentile in the distribution of effectiveness. These differences are quite large.

Comparatively, none of the differences for elementary teachers are statistically significantly different from 0. When we relax our threshold for statistical significance to allow for a 10% error/false positive rate, we do observe a difference for *knowledge of math pedagogy*. This relaxation seems reasonable given the smaller sample size and the substantively meaningful difference here of 0.37 SD, equivalent to moving the median teacher to the 64th percentile in the distribution of knowledge.

Figure 3: Teacher Ratings by Grade Level





Video Observations: Changes in Observed Teaching Practices

Several findings emerged across our analysis of the videotaped lessons. We present the themes that recurred across the video observations for both teachers. As a note, we are unable to make any judgments about teaching differences between teachers who participated in Activity 1 versus those who participated in Activity 2, as both teachers participated in both activities. Similarly, we are unable to offer a grade-level analysis, as both teachers taught elementary mathematics. To ensure confidentiality of the two teachers, we describe patterns in broad strokes rather than providing specific descriptions of the content and lessons.

When looking across the pre- versus post-implementation lessons for both teachers, we observed an increased push for student engagement. Both teachers asked more open-ended questions of their students in the second videotaped lesson compared to the first videotaped lesson. Additionally, each teacher encouraged more student participation in the second videotaped lesson compared to the first videotaped lesson. We also observed more examples of eliciting student thinking and encouragement of students to share their ideas to the class. Moreover, in the second lesson, there were more instances of the teacher c compared to the first lesson. In the second videotaped lesson, we observed less time that the teacher was talking and offering direct instruction and more time where students were talking or working independently.

The second theme that emerged across lessons for both teachers was a stronger use of technology and instructional materials. Both teachers relied heavily on instructional videos, but we observed more intentional pacing from both teachers in the second videotaped lesson compared to the first. For example, the teachers paused the videos more often to provide students time to work independently or pose questions.

We also observed growth from the first lesson to the second lesson with their use of manipulatives. It appeared that manipulatives were part of the curriculum for both teachers. At the same time, the teachers' use of manipulatives was stronger in the second lesson. Compared to the first videotaped lesson, in the second lesson the manipulatives were used to represent, explore and deepen student understanding. This included using the manipulatives to make linkages between multiple representations of a concept or to solve a problem using different solution strategies (e.g., multiplication using standard algorithm versus using unit blocks). Finally, the teachers gave students opportunities and encouraged them to use available physical and digital tools including fraction strips and instructional videos.

The final theme that emerged was some lack of precision in terms of definitions, notation, and vocabulary used to convey mathematical reasoning. Across both teachers and both lessons, we observed multiple instances of imprecise or incorrect mathematical language. For example, in one lesson, the teacher led students through a discussion of two types of area—"inside area" and "outside area"—whereas the materials suggested to us that the teacher wanted students to consider perimeter versus area. In another instance, when students were multiplying 300 by 2, the teacher stated that the "zeroes didn't matter," disregarding the importance of place value in the standard algorithm. Both are instances of imprecisions rather than mathematical errors where a problem is solved incorrectly. However, we note this theme, as it may be an area for continued professional learning. We also recognize that precision in mathematical language was not a key instructional focus area for either activity in the current professional learning series content.

Focus Group Discussions: Instantiation of Learning in Classrooms, and Feedback on Strengths and Weaknesses of Training

Of the five focus group discussions, one included a group that participated in Activity 1 only, two groups participated in Activity 2 only, a fourth group participated in both activities, and the fifth had a mix of individuals who engaged either in Activity 1 or Activity 2. One group included elementary instructors only, two groups included secondary teachers only, a fourth group included both elementary and secondary teachers, and the fifth group included district-level staff supporting teachers across a range of grade levels. Overall, themes that emerged from the focus group discussions did not differ substantially across activities or grade levels, though there are some exceptions that we call out in our discussion of themes below. Here, we have a larger sample of participants (compared to the videotaped lessons), and so we do provide specific illustrative examples and quotations.

Several common, cross-discussion themes emerged from the focus groups. First, participants shared their satisfaction with the content of the workshops. Broadly speaking, one teacher noted that "the content was phenomenal", and another noted that there was "lots of content to use." An administrator from Activity 2 particularly appreciated "the push for acceleration"; by helping to identify and work on essential standards that have to be taught (and learned) toward mastery, Teaching Lab gave a helpful "advanced look at what is coming next year" as this push is rolled out more directly from state-led initiatives.

More specifically, teachers and administrators described several ways in which they took up lessons from the training in their approach to and delivery of instruction. Across activities, teachers described how they had support to use manipulatives and hands-on activities in a way that fostered student engagement. One teacher noted how she was always "on the floor with students [using manipulatives] to help them learn concepts." Another teacher said that "the idea of having more visuals and real-world problems was helpful." While this was "not an area in which [she] is good," she knows "it will benefit my students."

Also observed across the two activities, teachers and administrators described how the focus on equity played out in their instruction. One teacher's comment nicely summarizes a common sentiment: with the support of Teaching Lab, she understands the importance of "making sure everyone is on the same plane. Not all students will use all the same resources. If one group is lacking formulas, [I] give them notes or an anchor chart." For other students, a different resource may be needed. Another teacher from this same focus group noted how the equity focus helped push her to really get to "know her students." She learned to "call on every kid", but also that "it is okay if not all students respond…; students have *opportunities* to talk to one another and to have productive struggles without leading to frustration."

Notably, when specifically asked to define "equitable math instruction", none of the focus group participants mentioned racial, ethnic, or cultural responsiveness. This pattern is consistent with our observations of instruction, where we observed changes in the extent to which teachers aimed to engage *all/as many* students as possible, but no specific changes in the other facet of equitable practices related to race, ethnicity, and culture.

A second theme that emerged across focus group discussions was an appreciation for and the benefit of cross-grade level and cross-school teams to engage in collaborative work. Having multiple opportunities to discuss their practice and learn from other educators—often from very different contexts—was salient. For example, one participant said, "analyzing how the instructional strategies have shifted, and being able to collaborate with someone on how to make those shifts helps." Collaboration amongst teachers was also highlighted by an administrator who said, "The use of different tools (jamboard, google docs, padlets, etc) helped. They gave [teachers] ideas on how to use the facilitation techniques to foster collaboration." In fact, several teachers and administrators—particularly from Activity 2—noted that they wanted even more collaborative time with schools in their network, and that they would be interested in continued support from Teaching Lab to engage in these sorts of professional learning communities into the future.

Cross-school and cross-grade level collaboration also helped with a push for coherence and vertical alignment. One teacher noted that "students don't get enough hands-on experience in math in the early grades, so when students come to her" it can be difficult to support student learning through manipulatives, etc. As such, the strong push from cross-grade collaboration allowed for greater understanding of how to create continuity in the student experience. One secondary school team described, in particular, the push for getting students ready for Algebra 1 and how the professional learning activities helped them think about vertical planning and use of student data to support this transition.

In terms of weaknesses, some participants expressed concern about some aspects of the facilitation of the workshops, but also cited that Teaching Lab facilitators were responsive to feedback. One teacher shared that the facilitators "took the feedback from us and made changes. For example, during the first session she read from a script, but by the end she improved". Several teams also described how the depth of group discussions was sometimes constrained by the lack of attendance and engagement. At the same time, they noted that low attendance was attributed to the Covid-19 pandemic and its impact on the transition to back-to-school learning. For some, cross-school teams never actually happened, because full schools dropped out before/right at the beginning of training. While lack of attendance and engagement was challenging for these individuals, they also recognized that Teaching Lab was limited in how they might better facilitate group discussions, when instances of low attendance were outside of their control.

Finally, across all focus group discussions, we identified some differences in tone and demeanor between groups of elementary versus secondary teachers/administrators. In particular, the elementary staff were more talkative and had more positive take-aways from the professional learning experiences. In contrast, the group of secondary educators had less to say about what was helpful or what could have been done differently. We recognize that these patterns differ from grade-specific patterns in our quantitative analyses. Thus, it may be that differences between elementary and secondary groups amongst the subset that agreed to participate in the focus group discussions may not generalize to all schools and teams.

Conclusion and Recommendations

We have grouped our recommendations into three broad areas: (1) benefits of working with Teaching Lab, (2) targeted improvements for the learning series, and (3) future research opportunities. We explore each area below.

Area 1: Benefits of Teaching Lab

Our findings suggest that teachers and school teams engaged meaningfully in the professional learning series—including both Activities 1 and 2—and that this engagement was associated with positive changes in desired outcomes. From these data, we draw two inferences:

- Additional schools and teachers in New Mexico could benefit from these experiences. Specific strengths of the program include the underlying content and materials, and opportunities for cross-school collaboration. In short, it may be beneficial to expand Teaching Lab's professional learning series to more districts, schools, and teachers.
- We heard from several teachers and school leaders that they would appreciate the opportunity to continue to have Teaching Lab's support to organize and facilitate cross-school collaborations, particularly as the state continues to roll out accelerated learning of the New Mexico Instructional Scope in the fall.

Area 2: Improvements to the Learning Series Content

Data captured primarily from the focus group discussions also reveal **ways in which the professional learning series content could be improved**:

- Particularly for Activity 2, teachers and school leaders noted that activities were meaningful but too short, both in terms of the number of sessions and the duration of each session. As such, Activity 2 might be lengthened**.**
- Largely because of Covid-19 and the return to in-person learning, creation of intact and fully participating cross-school teams was a challenge. Teaching Lab established cross-school teams based on a set of individuals who initially expressed interest; however, several of these individuals dropped out before/at the very beginning of implementation. As a result, some participants who stayed the course noted that their teams were small and lacked consistent engagement, which disrupted some of the learning. To help offset these challenges, future training may allow for a longer recruitment and planning period in order to fully gauge

prospective participants' interest and to ensure that attendance and engagement will be strong throughout.

- While participants, across the board, appreciated the Teaching Lab content and materials, some identified areas for improvement in the *delivery* of this content and the *facilitation* of sessions. In particular, several participants noted that they wished facilitation was not as scripted. At the same time, these same participants indicated that, when they provided this feedback to Teaching Lab, facilitation quickly changed in the following session.
- One of the stated goals of Teaching Lab's work is to address issues of equity in terms of *access* and *cultural responsiveness* in mathematics instruction. Participants spoke much more about the former than the latter. As such, it may be useful to increase the emphasis on culturally responsive mathematics instruction.
- In the observations of math lessons, we observed several instances of imprecise mathematical language. As such, future professional learning content may incorporate a focus on mathematical language and mathematical content knowledge, in addition to math pedagogy.

(3) Future Research Opportunities

Finally, we make **recommendations for continued research** that could be paired with ongoing (and revised) training to build on the findings described in this report:

- Additional research may seek to reconcile understandings about how engagement in professional learning activities translates into changes in desired outcomes. In particular, our analyses of the qualitative data revealed that elementary teachers and school teams, on average, were more "into" the activities than secondary teachers. Yet, quantitative analyses of pre/post differences in survey measures showed larger (positive) changes for secondary teachers than for elementary teachers. Thus, future research may probe school-level trends in greater detail, as well as the factors that may stand in the way of elementary (or other teachers) turning their interest and engagement in the content into changes in outcomes. In this report, we analyzed differences in attendance, finding similar patterns for elementary versus secondary teachers. Yet, attendance is an indirect proxy for engagement, and so future research may quantitatively explore measures of participant engagement more directly.
- To facilitate this sort of exploration, additional research will require larger sample sizes to support focused subgroup analyses. Subgroups analyses may examine differences between elementary and secondary teachers, as well as differences between activities and curricula used.
- Relatedly, there may be additional opportunities to examine substantive questions regarding future rollout of professional learning activities, including whether dosage makes a difference and how schools/districts can get the most "bang for their buck". Many focus group discussion participants—particularly from Activity 2—noted that they wished they had more time to dig into the content. We found somewhat larger changes in our survey constructs for Activity 1 versus Activity 2. At the same time, it is not appropriate in the current study to attribute these differences to dosage, as the two activities also differed in terms of content and materials. Therefore, future research could address the dosage question specifically by including subgroups that experience similar/the same content but with different numbers of facilitated sessions.
- Because the goal of this study was to understand the *effect* of Teaching Lab activities on desired outcomes, future research should compare individuals who participate in the professional learning series to those who do not. In other words, it would be beneficial to have a comparison group that did not experience any of the Teaching Lab activities in order to be

able to ensure that pre/post changes are not driven by natural improvements in practice over time. Further, to meet the gold standard of causal research, future research may seek to randomly assign individuals or school teams to Teaching Lab versus a control group. This approach ensures that findings are not driven by non-random selection into treatment. We recognize that schools and districts may hesitate to engage in a random-assignment study, as experiments often suggest that services will be withheld from teachers. However, we argue that random assignment could be done in a way that ensures that the needs of teachers and schools are met. For example, if there is greater demand than seats available, a study might randomly assign individuals to receive treatment in one year (or semester) versus to a control group who may receive treatment in the following year (or semester). Alternatively, the treatment group may receive the full Teaching Lab program, while the control gains access to the materials but not to the facilitated professional learning community sessions. Further, participants could be randomly assigned to receive more versus fewer overall sessions, as a means of testing the dosage question described above; in this case, *all* teachers would receive at least some baseline level of support.

- Finally, future research should consider capturing additional outcome measures that directly align with the goals of the professional learning series and of PED to ensure that all students have access to high-quality instruction and that student outcomes improve as a result. Specifically, future research may observe and score the quality of *all* teachers' instruction. In the current study, we observed instruction for a subset of teachers in order to qualitatively illustrate trends. In comparison, collecting rich classroom observation data from all teachers would facilitate quantitative analyses and to draw inferences to a broader population. Further, given that students are the ultimate beneficiary of teacher professional learning, it would be useful to collect and analyze student-level outcomes, whether that be test-score performance or student-reported experiences in the classroom.

References

Blazar, D., & Kraft, M. A. (2015). Exploring mechanisms of effective teacher coaching: A tale of two cohorts from a randomized experiment. *Educational evaluation and policy analysis*, *37*(4), 542-566.

Blazar, D., & Kraft, M. A. (2019). Balancing rigor, replication, and relevance: A case for multiple-cohort, longitudinal experiments. *AERA Open*, *5*(3), 2332858419876252.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.

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

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

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237-251.

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

Gay, G. (2002). Preparing for culturally responsive teaching. *Journal of Teacher Education*, *53*(2), 106-116.

Harris, D. N., & Sass, T. R. (2014). Skills, productivity and the evaluation of teacher performance. *Economics of Education Review*, *40*, 183-204.

Hill, H. C., Blazar, D., & Lynch, K. (2015). Resources for teaching: Examining personal and institutional predictors of high-quality instruction. *Aera open*, *1*(4), 2332858415617703.

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and instruction*, *26*(4), 430-511.

Hill, H., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review*, *83*(2), 371-384.

Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of labor Economics*, *26*(1), 101-136.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273-1296.

West, M. R., Kraft, M. A., Finn, A. S., Martin, R. E., Duckworth, A. L., Gabrieli, C. F., & Gabrieli, J. D. (2016). Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling. *Educational Evaluation and Policy Analysis*, *38*(1), 148-170.

Appendix 1: Survey Items

Appendix Table 1. Teacher and School Administrator/Principal Survey Items

Construct	Item Text
	Teacher Survey
To what extent do you agree or disagree with the following statements? (1 = Strongly Disagree to 5 = Strongly Agree)	
SWR	I take time to learn about the cultures represented by students in my classroom.
SWR	I communicate with parents of my students in positive ways, not just when there is a problem.
SWR	I know my students and build positive working relationships with them.
SWR	Creating a sense of community in my classroom is key to student success.
SWR	I build strong, positive working relationships with the parents of my students.
HE	It is not fair to ask students who are struggling with English to take on challenging academic assignments.
HE	Teachers should provide all students the opportunity to work on grade-level assignments and tasks.
HE	If I provide the proper scaffolds, all students in my class will be able to succeed with grade-level assignments and tasks.
HE	If a student doesn't learn something the first time, I will try another way.
HE	Teaching and learning that is aligned to math standards and to the curriculum is appropriate for developing all students' understanding of the subject area.
HE	The math standards and curriculum are too challenging for my students.
Please answer these questions based on your current teaching assignment. (1 = Strongly Disagree to 5 = Strongly Agree)	
SE	How well can you implement alternative teaching strategies to help students learn the content?
SE	To what extent can you use a variety of assessment strategies to help students learn the content?
SE	How much can you do to motivate students who show low interest in school work?
SE	How much can you do to get students to believe they can do well in school work?
SE	To what extent can you create standards-based learning opportunities that are engaging to your students?
SE	To what extent can you provide an alternative explanation or example when students are confused by the content of the lesson?
	Administrator/Principal Survey
Based on your knowledge of the teacher and their instruction, please rate them on the following characteristics. (1 = Needs Improvement to 7 = Exceptional)	
HQI	Overall, how effective is the teacher in delivering high-quality math instruction?
HQI	How effective is the teacher in supporting students to build conceptual understanding of math ideas?
HQI	How effective is the teacher in delivering math content that reaches all students?
HQI	How effective is the teacher in connecting with students as individuals and with their lives and cultures outside of the classroom?

Note: SWR = Strong Working Relationships; HE = High Expectations; SE = Self-Efficacy; HQI = High-Quality Instruction

Appendix 2: Video Observation Protocol

Domain	Sub-Domain	Specific Activities/Look-Fors	Comments
		Mathematics Practices	
Habits of Mind	Make sense of problems and persevere in solving them	• Involve students in rich problem--‐based tasks that encourage them to persevere in order to reach a solution • Provide opportunities for students to solve problems that have multiple solutions	
	Attend to precision	• Encourage students to focus on clarity of the definitions, notation, and vocabulary used to convey their reasoning • Encourage accuracy and efficiency in computation and problem-based solutions, expressing numerical answers, data, and/or measurements	
Reasoning and Explaining	Reason abstractly and quantitatively	• Facilitate opportunities for students to discuss or use representations to make sense of quantities and their relationships • Encourage the flexible use of properties of operations, objects, and solution strategies when solving problems • Provide opportunities for students to decontextualize (abstract a situation) and/or contextualize (identify referents for symbols involved) the mathematics they are learning	
	Construct viable arguments and critique the reasoning of others	• Provide and orchestrate opportunities for students to listen to the solution strategies of others, discuss alternative solutions, and defend their ideas • Provide prompts that encourage students to think critically about the mathematics they are learning	
Modeling and Using Tools	Model with mathematics	• Use mathematical models appropriate for the focus of the lesson • Encourage student use of developmentally and content--‐ appropriate mathematical models (e.g., variables, equations, coordinate grids) • Remind students that a mathematical model used to represent a problem's solution is 'a work in progress,' and may be revised as needed	
	Use appropriate tools strategically	• Use appropriate physical and/or digital tools to represent, explore and deepen student understanding • Teacher gives student opportunities and encourages students to use available physical and digital tools	
Seeing Structure and Generalizing	Look for and make use of structure	• Engage students in discussions emphasizing relationships between particular topics within a content domain or across content domains • Provide activities in which students demonstrate their	

		flexibility in representing mathematics in a number of ways e.g., 76 = (7 x 10) + 6; discussing types of quadrilaterals, etc.	
	Look for and express regularity in repeated reasoning	• Engage students in discussion related to repeated reasoning that may occur in a problem's solution • Draw attention to the prerequisite steps necessary to consider when solving a problem • Urge students to continually evaluate the reasonableness of their results	
<u>Equity Practices</u>			
Affirming Mathematics Learners' Identities	Affirming different racial, ethnic, and cultural identities as related to mathematics	• Showing diverse representations of mathematicians and scientists and/or historical and current methods/strategies prominent in other cultures and countries (e.g. Chinese methods of foiling) • Evidence that the teacher makes sensitive use of examples. • Teacher demonstrates a working knowledge of the cultures and identities of the students in their class.	
	Bridging gap between formal and informal experiences as related to mathematics	• Validating student experiences with math inside and outside of the classroom • Students are often asked to apply the math they learn to the world around them.	
	Student-oriented classrooms (giving students power in the classroom)	• Highlighting and centering students' methods of solving problems as valuable • Evidence of the teacher encouraging and supporting a diverse array of mathematical competence. • Affirms and draws on students of knowledge.	
Challenging spaces of marginality	NA	• Evidence of the teacher deliberately expressing (high) expectations that everyone is able to do the work. • Teacher uses a variety of teaching practices to foster a communal learning environment (e.g., collaboration, group work, distributing talk time amongst students and teacher) • Teacher encourages multiple types of participation (verbal, written, kinesthetic) for supporting students to demonstrate their mathematical competence • Teacher encourages students to use more than just traditional algorithms to solve problems • Some of the teacher's questions have known/correct answers, but many encourage mathematical thinking. • Note who is doing the talking/teaching	

Appendix 3: Focus Group Discussion Protocol

Thank you everyone for joining us today. My name is [insert researcher's name] and I am [insert researcher's title]. Our goal today is to hear from you about your experiences with the Teaching Lab Professional Learning Series and to provide feedback where relevant. Our conversation will last no more than 1 hour.

Before we dive into the conversation, we need indication that you agree to take part in this discussion for research purposes. Some of you sent in signed consent forms prior to today. For those who were not able to do so, I am putting a link to the consent form in the chat. Can you each take a couple of minutes to review that, electronically sign, and either email back to me or send back through the chat?

To begin our conversation, let's have everyone state their name, position at their school, and their years of experience in that role.

1. In general, how was your experience with Teaching Lab's professional learning? In what ways did it help you as a teacher or school leaders?
2. [GETTING AT CHANGES IN MINDSETS]: In thinking about mathematics instruction, how would you define "equity"? How has this definition changed (if at all) since working with Teaching Lab?
3. [GETTING AT CHANGES IN PRACTICES]: What were some of the practices in your classroom you changed as a result of working with Teaching Lab?
    a. What are some examples of "high-leverage" practices you <u>now</u> use with your students?
    b. What are some examples of *equitable* practices in math instruction you <u>now</u> use with your students?
    c. [FOR ACTIVITY 1 TEACHERS]: In what ways has work with the curriculum supported implementation of these practices?
    d. [FOR ACTIVITY 2 TEACHERS]: What are some examples of practices you <u>now</u> use to accelerate student learning through utilization of the New Mexico Instructional Scope?
4. [GETTING AT TEACHING LAB MECHANISMS THAT SUPPORT CHANGES IN MINDSETS AND PRACTICE]: What were some of the practices that Teaching Lab facilitators used in the professional learning series that were most helpful in supporting changes in your instruction?
    a. What are some of the ways the Teaching Lab has empowered you to support your students?
    b. How has Teaching Lab's professional learning helped you with your teaching or support for students during the <u>Covid-19</u> pandemic?
5. What suggestions or recommendations would you have for Teaching Lab to improve their professional learning in the future?

6. Finally, is there anything I haven't asked you that I should have? If we have any additional questions or need clarification on any of the points that were made today, may we contact you?

Thank you very much for helping us out today. Your feedback will be very useful to us. It is all right to talk to others about what we discussed here today, but please remember to respect each other's privacy, and don't mention anyone's name outside this room.