

Leveraging Mathematics Teacher Leaders in Support of Student and Teacher Learning

Nicole Rigelman & Chandra Lewis

To cite this article: Nicole Rigelman & Chandra Lewis (2022): Leveraging Mathematics Teacher Leaders in Support of Student and Teacher Learning, *Investigations in Mathematics Learning*, DOI: [10.1080/19477503.2022.2140989](https://doi.org/10.1080/19477503.2022.2140989)

To link to this article: <https://doi.org/10.1080/19477503.2022.2140989>



Published online: 31 Oct 2022.



Submit your article to this journal [↗](#)



Article views: 45



View related articles [↗](#)



View Crossmark data [↗](#)



Leveraging Mathematics Teacher Leaders in Support of Student and Teacher Learning

Nicole Rigelman ^a and Chandra Lewis^b

^aDepartment of Curriculum and Instruction, Portland State University, Portland, Oregon, USA; ^bRMC Research Corporation, Portland, Oregon, USA

ABSTRACT

Transforming mathematics learning and teaching toward more equitable and effective approaches is critical to student mathematics learning and identity development. This task at-scale in a district takes time, commitment, and mathematics expertise that may not be widespread in the absence of focused professional development. District and regional mathematics leaders with university mathematics and mathematics education faculty partnered to address this challenge by designing a professional development program that prepared K-12 teachers of mathematics as leaders at the classroom, school, and district level. Described are specific aspects of a professional learning model focused on developing mathematics content, pedagogical, and leadership knowledge and skills. Also provided are findings related to the impact of the project's professional development on shifts in instructional practice and student achievement.

KEYWORDS

Mathematics specialists; teacher leaders; coaches; professional development; teacher learning; student learning

Introduction

As schools and districts aspire to transform the teaching of mathematics to be problem-based and discourse-rich, many turn to using mathematics specialists to support instructional shifts (National Council of Teachers of Mathematics, 2014). Because having a mathematics specialist who serves as a coach, working with adults, is an added expense for schools and often removes highly skilled teachers from their work with students, examining the influence and impact of both mathematics coaches and classroom-based teacher leaders is critical. An NCTM research brief detailed 24 research studies of mathematics coaches (McGatha et al., 2015). Across the studies, researchers saw improvements in teaching practices connected to the focus of the professional learning (e.g., questioning, mathematical discourse, student engagement, conceptual understanding, formative assessment); however, there were only six studies that examined student achievement. At present, there is minimal research on how mathematics specialists that remain at least partially in the classroom can support improved teaching practice across a school and on how both types of specialists can positively influence student achievement.

An examination of the student achievement literature pertaining to mathematics specialists (MSs) revealed that almost all the studies examined student achievement at either the elementary or middle grades, and most of the research focuses on the impact of mathematics specialists in a coaching role (Balfanz et al., 2006; Brosnan & Erchick, 2010; Campbell et al., 2017; Campbell & Malkus, 2013; Coniam, 2010) rather than as classroom teachers (Meyers & Harris, 2008; Nickerson, 2010; Zollinger et al., 2010). When studying the impact of mathematics specialists on student achievement, the research designs typically included either school level (Brosnan & Erchick, 2010; Nickerson, 2010; Zollinger et al., 2010) or student level data on state administered assessments (Balfanz et al., 2006; Campbell et al., 2017; Coniam, 2010; Meyers & Harris, 2008). Of the six studies that included student level data, one was limited by

a small student sample size (Coniam, 2010), another was unable to obtain true baseline data (Campbell et al., 2017), and only one reported specifically on closing the achievement gap among subgroups (Balfanz, MacIver, & Byrnes, 2006). One study (Meyers & Harris, 2008) revealed that students who were in schools with more than one mathematics specialist experienced significantly greater gains than students in schools with only one mathematics specialist. Studies by Campbell and Malkus (2013) & (2017)) indicated that increases in student achievement were often not seen within a year of treatment but that several years were needed to see a significant impact on student achievement.

The present study focuses on the preparation, influence, and impact of well-prepared mathematics teacher leaders (MTLs) by providing evidence of their content and pedagogical knowledge growth, their use of research-proven teaching practices in their classrooms, and the achievement of their students. The 3-year East Metro Mathematics Leadership (EaMML) Math-Science Partnership project was a collaboration among David Douglas School District (District 1), Centennial School District (District 2), the Multnomah Education Service District, Portland State University, and RMC Research. The EaMML districts sought to develop district-wide mathematics leadership teams (MLTeams) representing every school in the districts, and including individuals serving a variety of roles: teachers, elementary instructional coaches, principals, and district mathematics leaders. These teams would engage in deepening their mathematics content, pedagogical, and leadership knowledge and skills to better serve students in their schools. A subset of the MLTeam members were also mathematics leadership cadre (MLC) members who, in addition to leading through their roles as teachers, coaches, or administrators, were charged with disseminating their learning by providing professional learning for non-EaMML teachers at the grade, course, school, or district level.

There are several unique features of the EaMML project that add to the research base. One unique feature of the EaMML project is the K-12 representation of math specialists engaging in both across-the-grades and within-grade-band professional learning. The math specialists included both formal (i.e., coaches) and informal leaders (i.e., elementary teachers teaching all subjects as well as middle- and high-school mathematics teachers). A second unique feature is studying the preparation and positioning of classroom teachers as mathematics leaders, as most of the research focuses on mathematics specialists as coaches. Few studies examine the impact of classroom-based mathematics specialists on student achievement. Not only did the EaMML project utilize student-level state assessment data in a rigorous quasi-experimental design, but the project also examined data across elementary and middle school (grades 3–7) and from an equity perspective, investigating if student achievement gains were experienced consistently across gender, race, ethnicity, and socio-economic status. Finally, to provide more information about translation of teacher learning to instructional practice, the EaMML project collected pre-post, K-12 teacher-level artifact and observation data. While several studies include observation data, there were no studies that also included artifacts.

Research Questions

EaMML's research questions focus on the project's impact on both the MTLs and their students.

RQ1: What was the influence of EaMML professional development on MTL mathematical content knowledge, pedagogical knowledge, and leadership knowledge and skills?

RQ2: To what extent has the EaMML project increased the school and district capacity to provide mathematics professional development?

RQ3: To what extent did EaMML MTLs use research-proven instructional practices that develop deeper student understanding of mathematics in their classrooms?

RQ4: Did students who had an EaMML MTL demonstrate higher achievement on state assessments than students who did not have an EaMML MTL?

Table 1. EaMML leadership types.

Leadership Type	Description
Project Leadership (<i>n</i> = 8)	Individuals from each partner who guided project development and supported implementation through their facilitation of various professional learning components.
Mathematics Leadership Team (MLT) (<i>n</i> = 94)	Individuals from every school in the districts were identified to serve as the MLT and engaged in ongoing learning through project professional development (described in upcoming section). This group included teachers, elementary instructional coaches, principals, and district mathematics leaders. They led by supporting effective and equitable mathematics teaching through their role (i.e., modeling effective and equitable practices, actively engaging in mathematics professional learning communities serving as a “more knowledgeable other”).
Mathematics Leadership Cadre (MLC) (<i>n</i> = 29)	Beginning in year 2, 29 of the 94 MLT members volunteered to serve as part of the MLC. This included teachers, elementary instructional coaches, and district mathematics leaders. As informal or formal leaders, MLC members led by designing and facilitating professional learning for their grade-level or course team, school, and district colleagues.

Note: The *n*'s reflect the total number of recruited participants.

Definition of Mathematics Teacher Leaders for the EaMML Project

Table 1 provides a description of the three leadership participant types for the EaMML project. EaMML focused on two types of leaders—those who participated through the MLTeam and those who also served through the MLC. While the project focused on these two types of leaders holding varied roles (i.e., teachers, instructional coaches, administrators), this manuscript focuses on the teaching and learning outcomes for the EaMML MTLs; specifically, these are the MLTeam's classroom-based teachers who completed the 3-year project (i.e., 51 of the 67 teachers), of which 19 were also part of the MLC.

A potentially surprising characteristic of the MLC was that formal leadership roles did not guarantee engagement with providing school- and district-level mathematics professional development for the EaMML project. For example, of the 14 elementary instructional coaches in the project, only three chose to participate in the cadre. The vast majority (i.e., 76%) of the MLC members were classroom teachers, this in contrast with having the majority in formal leadership roles (i.e., school-level coach, district-level specialist, or administrator). This is similar to the overall relationship between informal and formal leaders among the MLTeam members, suggesting that EaMML mathematics teacher leader participants were not only leading from within their classroom but also outside their classrooms through their role on the MLC.

Conceptual Framework for the Preparation of EaMML Mathematics Teacher Leaders

NCTM's *Principles to Actions* (NCTM, 2014) was used to develop a shared vision for effective and equitable mathematics teaching among project leaders and participants while providing a foundation for the various professional development (PD) components. Central to all the PD components was the emphasis on implementation of high cognitive demand tasks, connected representations (i.e., physical, visual, verbal, symbolic, and contextual), and meaningful student mathematical discourse in support of deep mathematics learning, which provided coherence across the PD components.

EaMML's desired outcomes included both deepening teachers' professional knowledge and skills as well as increasing the district and school capacity for providing effective mathematics learning experiences for teachers and students. Because of this, EaMML's PD model included a combination of job-embedded professional development approaches for all EaMML participants (i.e., book studies, lesson studies, curriculum mapping, assessment development) and the option to participate in standards-aligned graduate-level mathematics specialist (MS) coursework (Association of Mathematics Teacher Educators, 2013; NCTM, 2014; Sutton et al., 2011). Summarized in Table 2 are the various components of the EaMML Professional Development program, the total number of hours, the facilitators, and the participating groups (i.e., MLTeam, MLC, other teachers).

Table 2. EaMML professional development activities.

Activity Description	Total Hours	Facilitators*		Participants		
		Project Leaders	MLC	MLT	MLC	Others
Kick Off Events. Beginning in Spring 2014 and again in Fall 2015 and 2016, teacher, coach, and administrator participants in each district engaged in daylong events focused on connecting EaMML professional learning with research-proven instructional and leadership practices. In Fall 2017, there was a project wide celebration that included a presentation and celebration of the research results (i.e., EaMML effect on teacher and student learning) as well as planning for continued work following the project.	28			•	•	
Book Studies. Book studies, offered 3 times in Spring 2014 and 4 times in each subsequent year for 2 hours, supported development of teacher and coach knowledge about pedagogy and leadership; the first read, <i>Principles to Actions</i> (NCTM, 2014), supported development of a shared vision for effective and equitable mathematics teaching and learning. Additional books were selected to help teachers and coaches develop the skills (a) eliciting and using student mathematical thinking to inform instruction, (b) designing/adapting tasks to be responsive to and relevant for students, and (c) honing leadership skills.	14		s	•	•	
Lesson Studies. Lesson studies, offered Spring 2014 and 3 times in two subsequent years to grade band teams, supported development of specialized mathematical content and pedagogical knowledge. The groups convened in a lead team member's classroom for a 1-day modified lesson study (Hurd & Lewis, 2011; Watanabe, 2002). Instead of <i>tuning a lesson</i> , this work focused on <i>tuning practice</i> through collaboratively “doing the math,” planning for student engagement in high-level thinking and discourse, observing for student mathematical discourse, debriefing, and planning next steps—all keeping effective and equitable teaching practices at the fore.	49		s	•	•	
District-Based Curriculum Work. Each summer, district mathematics specialists convened teachers and coaches to work on a combination of activities that included: establishing a mathematics vision, engaging in curriculum review, developing curriculum and assessment maps and associated resources. Common features across the collaborative work were examination of standards and progressions, analysis of tasks and curriculum materials including assessments to identify and address potential gaps.	28			•	•	
Optional:** Mathematics Specialist Coursework. EaMML teachers and coaches were invited to participate in graduate-level courses aligned with AMTE's <i>EMS Standards</i> (AMTE, 2013). These courses were clustered into content-focused pedagogy and leadership courses—seven 30-hour courses in all. Focal content for the included Base Ten Numeration and Operations; Whole Number and Fraction Operations; Generalizations about Operations; Patterns and Functions; Measurement; Data Analysis where DMI materials served as the foundation. Participants completed 5 of the 6. The leadership course and practicum focused on leading both within and outside the classroom.	210	, s, +		•	•	•
Building- and District-Based Professional Learning. Beginning in Fall 2016, Mathematics Leadership Cadre Members began supporting project-sponsored professional learning such as the book studies and lesson studies. They also offered professional learning through coaching and mentoring colleagues individually or through their grade-level or course teams as well as leading colleagues in book studies or lesson studies.	169	s			•	

*| = led, s = supported, + = included other PSU facultyHT

**Mathematics Specialist Courses were completed by a subset of MLT and MLC members. The courses were also open to district colleagues.

The purpose of selecting job-embedded PD approaches not only supported careful study and transformation of day-to-day instructional practice (Ball & Cohen, 1999; Smith, 2001; Sztajn et al., 2017), but it also allowed MLC members to replicate these structures and content as they shared learning with their colleagues. Similarly, the MS coursework centered on learning from practice with expectations to implement course-based learning when working with their students and colleagues. The intended by-product of the project was to cultivate leaders who would help sustain ongoing mathematics PD for all district teachers.

Garet et al. (2001) identified three structural features of effective professional development as (1) the *form* of the activity, (2) the *duration* of the activity, and (3) the degree to which the activity promotes *collective participation*. These structural features along with the unique core features defined by Desimone (2009)—content focus, active learning, coherence—were considered in the EaMML PD design. The PD model included six components described in Table 2. All the PD components reflect reform-type PD. On the surface, the MS Coursework might be considered traditional PD but the case-based curriculum design (i.e., written and video cases, cases of one’s own students) supports teacher leaders with developing deepened mathematical understanding simultaneous to developing pedagogies responsive to their students’ thinking. Four of the six activities included all participants (i.e., kickoff events, book studies, lesson studies, and district-based curriculum work) with a subset of participants—specifically 20—completing four or more courses in the mathematics specialist program and 29 engaged as MLC members, leading professional learning within their team, school, or district. Taken together, all EaMML participants engaged in sustained professional learning logging more than 50 hours during each of the full two years. The 20 participants engaged in coursework had an additional 90 contact hours per year. MLC members—not including district leaders—led an additional 169 hours of professional learning in the second full year. Across the professional learning components teachers engaged in purposeful collaboration both within-grade or across-the-grades focused squarely on transforming students’ opportunities to learn.

The desired pedagogical outcomes from the EaMML professional learning were to enhance participants’ professional knowledge and skills so they could make shifts in instructional practice that would positively impact students’ mathematics learning. Each component of the EaMML PD supported participants with deepening their understanding of the content they taught and the ways students learn that content. Through participation in the lesson studies, teachers and coaches put their learning from the book studies and the courses into action while they also developed skills with noticing, analyzing, and responding to students’ thinking. As described in Doerr et al. (2010) a focus on student thinking and classroom practice can lead to developing productive habits of mind focused on continuously learning from teaching.

The desired leadership outcomes from the EaMML professional learning were to prepare participants for their leadership role by broadening their perspectives about leadership and equipping them to support collaborative professional learning. Instead of viewing leadership as an individual, EaMML project leaders wanted participants to view leadership as a shared practice (Harris, 2003; Muijs & Harris, 2003). This view of shared or distributed leadership positioned teacher leaders to lead formally and informally, sharing their expertise both within and beyond their classrooms (c.f. Wasley, 1991). EaMML participants had the opportunity to deepen their understanding of how adults learn and models for supporting that learning while simultaneously developing their identity as a teacher and leader of mathematics. Through participation in the book studies and lesson studies, teachers and coaches experienced learning in the ways project leaders hypothesized could be used to support continued learning at the grade, course, school, and district levels. Teachers and coaches participating in the MS coursework also completed field-based projects where they engaged in a coaching cycle supporting a new or experienced teacher and facilitated professional learning at their grade or school level. In the last year of the grant, MLC members not only supported facilitation of grant sponsored events but also led learning for their team, school, district, and beyond.

Methods

Participants

The project was a 3-year collaboration from 2014 to 2017 among one large city school district (14 schools, approximately 9,800 students, 75% free and reduced lunch (FRL), and 55% Black and Indigenous People of Color (BIPOC)), a large suburban district (9 schools, approximately 6,000 students, 77% FRL, and 60% BIPOC), an education service district, a university, and a research

firm. In the 2016–2017 school year, approximately 75% of the students in both districts had access to free and reduced lunch, and 55% of the students were BIPOC in District 1 and 60% in District 2. Of the 94 recruited EaMML MLTeam members, 67 were classroom teachers, and 51 completed the 3-year project and provided both pre and post data, hereafter referred to as EaMML teachers. Of the 51 EaMML teachers, four taught Grade K, four taught Grade 1, three taught Grade 2, six taught Grade 3, four taught Grade 4, one taught Grades 4 and 5, five taught Grade 5, six taught Grade 6, one taught Grade 7, two taught Grade 8, four taught Grades 6–8, and eleven taught a range of mathematics courses in Grades 9–12.

Research Question 1: Mathematics Teacher Leader Learning

Design

Research Question 1 is addressed using two pretest-posttest designs. EaMML teachers completed a survey at the onset of the project (pretest) and at the conclusion of the project (posttest). The matched pre-post completion rate was 76% (51 of 67). The research team developed the survey to measure changes in teachers' pedagogical and leadership knowledge and skills. The research team reviewed a 121-item teacher survey developed by the Arizona Mathematics Partnership (AMP), a 5-year Mathematics and Science Partnership project funded by the National Science Foundation (Weaver et al., 2018). The research team selected items from the AMP survey, modified some wording of the items, and developed items that aligned with the instruments for the other designs. The survey includes one scale measuring leadership with three items that directly aligned with the definition of leadership for this project. These focus on teacher collaboration, sharing instructional materials, and leading professional learning. There were six scales measuring different types of pedagogical knowledge: (1) addressing *Common Core State Standards in Mathematics*, (2) developing mathematical tasks, (3) addressing students' learning needs, (4) assisting students in sense making, (5) encouraging student discourse, and (6) developing students' mathematical reasoning. The research team calculated the Cronbach's alpha score for each scale to determine the reliability and internal consistency of each scale. The Cronbach's alpha scores for all scales exceeded 0.67.

For the second pre-post design, EaMML teachers completed one of the Mathematical Knowledge for Teaching (MKT) assessments (Hill et al., 2004) at the beginning of the project (pretest) and at the end of the project (posttest). The matched pre-post completion rate was 76% (51 of 67). The MKT measures changes in participants' specialized knowledge for teaching mathematics. This project used the Number Concepts and Operations (elementary school and middle school versions with elementary and middle school participants respectively) and the Patterns, Functions, and Algebra (with high-school participants) assessments, because these assess topics addressed in the EaMML professional development. Utilizing Item Response Theory, the MKT assessment scales have reliabilities between 0.75 and 0.85.

Analysis

Paired t-tests were used to assess differences between pre and post. Differences were deemed statistically significant if $p < .05$. Significance tests were only conducted at the aggregate level; however, the data were broken out by key subgroups to determine descriptive differences between MLC and MLTeam participants, those teaching Grade K–5 and Grades 6–12, and those who completed the MS coursework through the practicum (i.e., 24 quarter credits of graduate-level coursework) and those who had not. Significance tests were not conducted for each subgroup due to limited sample sizes and to reduce the probability of a false-positive finding.

Table 3. EaMML survey pre and post data (Means) by subgroup.

	<i>n</i>	Instructional Leadership*		Common Core State Standards*		Mathematical Tasks*		Learning Needs*		Sense Making*		Discourse*		Reasoning*	
		Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Overall	51	3.29	3.52	3.22	3.67	2.91	3.53*	3.31	3.55	3.15	3.55	2.97	3.45	2.74	3.34
Grade K-5	27	3.17	3.53	3.07	3.69	2.74	3.47	3.12	3.44	2.96	3.56	2.82	3.47	2.54	3.27
Grade 6–12	24	3.43	3.50	3.38	3.65	3.10	3.60	3.51	3.67	3.37	3.53	3.14	3.42	2.96	3.42
Practicum	10	3.50	3.83	3.25	3.90	2.97	3.67	3.00	3.50	3.08	3.78	2.95	3.80	2.70	3.50
MLT	32	3.20	3.47	3.17	3.64	2.86	3.56	3.28	3.49	3.08	3.49	2.90	3.42	2.72	3.32
MLC	19	3.46	3.60	3.29	3.71	2.98	3.47	3.35	3.65	3.27	3.65	3.09	3.49	2.77	3.37

Survey Response Options: 1 = *Not at all*, 2 = *A little*, 3 = *Somewhat*, 4 = *Very*. Only matched pre and post data are included. Paired *t*-tests used to assess differences between pre and post at the aggregate level. *Differences deemed significant if $p < 0.05$

Results

The results indicate that EaMML teachers ($n = 51$) increased their leadership knowledge and pedagogical knowledge after completing the EaMML professional development program. EaMML teachers' scores on all seven four-point scales increased significantly over time (see Table 3). The average post scores were all high (i.e., three or higher). An examination of subgroup differences revealed that those who completed the practicum had the highest post scores ($M = 3.83$) along with MLC members ($M = 3.60$) for leadership knowledge. For the pedagogical knowledge scales, those who completed the MS coursework through the practicum had the highest post score for five of the six scales and MLC members' post scores were higher than the MLTeam members' post scores for five of the six scales.

Closer examination of the items that comprise the leadership scale indicated which aspects of leadership were strongest for MTLs. Two of the survey items increased significantly from pre to post: “talk about math teaching and learning with colleagues” (Pre, $M = 3.55$; Post, $M = 3.73$) and “design and lead professional learning sessions for peers” (Pre, $M = 2.69$; Post $M = 3.06$). The item “share instructional materials with colleagues” did not increase significantly from pre to post (Pre, $M = 3.65$; Post, $M = 3.75$); however, teachers already felt confident in this area at the time of the Pre. Although MTLs significantly increased in terms of designing and leading professional learning for their peers, this was the area in which they rated themselves the lowest.

The MKT results indicate that EaMML teachers increased their mathematical and pedagogical knowledge; scores improved on all assessments and elementary and high-school participants' scores increased significantly (see Table 4). An examination of subgroup differences reveals that the MLC members' post scores were higher than the MLTeam members' post scores on all assessments. The results for the mathematics leadership practicum subgroup were mixed: those completing the elementary version of the assessment obtained the highest post score; however, those completing the middle school version had the lowest post score.

Research Question 2: Increased Capacity

Design

Research Question 2 is addressed by collecting data on the mathematics leadership support MLC members provided to their colleagues. In the 2016–2017 school year, the research team documented the number of professional development events, the number of hours, and the specific audience for each event (i.e., grade level or course team, school-based, district-based, and outside the district) provided by the MLC teacher leaders. Because expectations for leading professional learning for colleagues were late in the EaMML project, the research team collected data a year and a half after the conclusion for the grant by surveying MLC members (18 of the 19 were still at the district) and interviewing the remaining district mathematics specialist to determine which, if any, activities

Table 4. Mathematical knowledge for teaching pre and post results by subgroup.

	<i>n</i>	Pre	Post
Number Concepts and Operations (ES)*	30	0.3793	0.9315
Practicum*	8	0.8196	1.7798
MLC	11	0.6251	1.1064
MLT	19	0.2370	0.8303
Number Concepts and Operations (MS)	12	0.5604	0.9646
Practicum	2	0.5907	0.0892
MLC	6	1.1922	1.3630
MLT	6	−0.0715	0.5662
Patterns, Functions, and Algebra (HS)*	9	0.9249	1.5614
MLC	2	0.7178	1.9500
MLT	7	0.9841	1.4503

Only matched pre and post data are included. $n = 30$ for NCOP (ES). $n = 12$ for NCOP (MS). $n = 9$ for Patterns, Functions, and Algebra (HS). Scale: NCOP ES ranges from -2.75 to 3.02 ; NCOP MS ranges from -2.80 to 2.53 ; Patterns, Functions, and Algebra ranges from -3.35 to 2.19 . Paired t tests were used to determine whether aggregate gains were significant. *Differences were deemed significant if $p < 0.05$.

continued, how often, and the supports and hindrances faced as they continued the mathematics-focused work.

Analysis

The research team used the project records to report on the frequency and duration of professional learning events. The follow-up survey data were summarized using descriptive statistics and open-ended items were analyzed for themes. The follow-up interview was analyzed to identify themes and used to better understand the contexts influencing the follow-up survey results.

Results

Across the two districts, there were 98 events led by MLC teacher leaders in the 2016–2017 school year, 60 in District 1 and 38 in District 2, for a total of 169 hours, 116 hours in District 1 and 53 hours in District 2. [Figure 1](#) shows the range of professional learning activities including by grade level, school based, district based, and outside of the district.

The follow-up survey results revealed the positive impact of the project-purchased lending libraries for book studies and protocols for analyzing student work or students at work such as in lesson study. These resources supported MLC members to lead book studies (41%), lesson studies (59%), or aspects of lesson study (59%) and were the most common ways the MLC teacher leaders continued to lead once the project ended. These MTLs, with the support of their grant-inspired school-based mathematics leadership teams and MTL roles (i.e., mathematics resource teachers in District 1, and mathematics residents in District 2), regularly facilitated or co-facilitated learning for their teams, schools, and districts (41%). Fifty-nine percent reported leading professional learning on a weekly/bi-weekly basis, 35% on a monthly/quarterly basis, and 6% reported only annual opportunities to lead professional learning. In the interview, the district mathematics specialist reflected “we are seeing teachers opt in to more professional learning and begin to assist with the planning and facilitation of district professional learning. Those who were part of EaMML are often quicker to step into the lead facilitator role than those who have not had as much training.”

MLC teacher leaders also revealed supports and hindrances to improving mathematics teaching and learning. Following EaMML, each district created MTL roles. Most saw these leadership opportunities as a lasting success of EaMML, with 76% of the MLC members serving in these roles. Others reflected positively on opportunities to influence their colleagues’ instructional practices through opening their classrooms for colleagues to visit and collaborating within their teams. The most

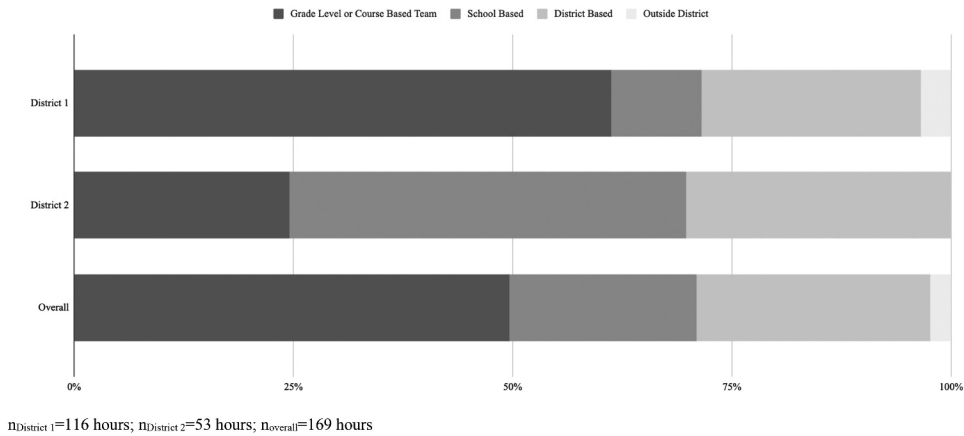


Figure 1. Percent of professional development activities by audience.

common challenges identified were skepticism or resistance from colleagues (~24%) and lack of time to collaborate (19%). Finally, a hindrance for some was also the way MTLs are positioned as voluntary or supplemental rather than central supports to leading change (i.e., math focused coaches, compensated roles).

Research Question 3: Mathematics Teacher Leader Instructional Practices

Design

Research Question 3 was addressed using a mixed-method approach utilizing both artifacts and observations. Both were scored using the Instructional Quality Assessment (IQA; M. Boston & Wolf, 2004) which focused on Academic Rigor in both tools, and added Accountable Talk, and Accountability to Knowledge and Rigorous Thinking for the observations. These domains are aligned with focus areas in the EaMML professional development. Artifacts were scored using the three academic rigor rubrics: Potential of the Task, Implementation of the Task, and Teacher Expectations. Each domain on the artifact rubric could receive a slightly different score range; Potential of the Task had a range of 0 to 4, Implementation of the task 1 to 4, and Teacher Expectations, 1 to 4 and not applicable. For the observation rubrics the following domains could receive a score of 0 to 4: Potential of the Task, Implementation of the Task, Student Discussion Following the Task, and Mathematical Residue. The Questioning, Participation, Teacher Linking, Student Linking, Asking, and Providing domains could receive a score of 0 to 4 or not applicable. EaMML project leadership deemed scores of 3 and 4 as reflective of optimal teaching practices. Scores 3 and 4 focused on, for example, high cognitive demand tasks and conceptual understanding rather than a focus on procedural skill; 3 s and 4 s focused on questions pressing for justification of why versus telling how; and 3 s and 4 s focused on sharing and comparing multiple strategies rather than privileging a single solution path.

EaMML teachers submitted classroom artifacts at the onset of the project (pre) and at the conclusion of the project (post). The matched pre-post completion rate was 75% (50 of 67 teachers submitted artifacts at the onset and completion of the project). The classroom artifacts included (a) a cover sheet for the teacher to record general information, (b) a mathematics task that either the teacher developed or borrowed from a curriculum resource, (c) the rubric the teacher used to score the student work, and (d) four examples of student work (i.e., work receiving a low score, middle score, and high score, as well as an example the teacher found interesting).

The research team designed an objective and rigorous artifact scoring training. The artifact scoring process involved each member of EaMML project leadership team (a) reviewing the artifact and cover sheet, the student task, the rubric the teacher used to score the student task, and examples of student work; (b) scoring the artifact independently using IQA Rubrics, (c) discussing the individual IQA scores and discrepant scores, and (d) determining and recording a final consensus score for data analysis. The research team calculated pre and post interrater reliability. Reliability was at least 0.67 before coders conferred.

For the observations, a random stratified sample of 31 EaMML teachers were selected for pre and post observation. Of the 31 teachers, 26 were observed in Year 1 (pre) and Year 3 (post) for a pretest-posttest design. As with the artifacts, the research team used the IQA Classroom Observation Rubrics due to their alignment to the project's professional learning goals and artifact rubrics.

The observation training and scoring processes were similar to those used for the artifacts. The ten coders met multiple times in Fall 2015 to view classroom videos and code videos in an effort to increase reliability between coders prior to conducting the baseline observations. During the training, and through the duration of the project, the research team conducted analyses to assess interrater reliability among observers. Overall, the reliability of the observers' independent scores was high (i.e., at least 0.73); however, there were some domains with lower reliability. Therefore, there were always two observers at each observation, each observer independently recorded scores, and then discussed and arrived at consensus scores for each domain.

Analysis

Independent *t*-tests and Mann–Whitney U-tests were used for between-group comparisons, paired *t*-tests, and Wilcoxon tests were used for within group comparisons. For all tests, differences were deemed statistically significant if $p < .05$. As with the prior research question, significance tests were only conducted at the aggregate level and key subgroup differences are provided descriptively.

Results

The results indicate that EaMML teachers increased their use of research-proven instructional practices after completing the EaMML professional development program. Artifact data ($n = 50$) indicate that teachers increased their use of research-proven instructional practices that develop deeper student mathematics understanding: scores increased over time in all areas and significantly for two of the three areas as shown in Table 5. EaMML teachers' post scores were highest for Potential of the Task. None of the average post scores were within the range of strong mathematical practices (i.e., a score of 3 or higher).

The research team examined key subgroups to determine if there were differences. An examination of subgroup differences reveals that the MLC members' post scores were higher than the MLTeam members' post scores for all three components, those teaching Grade K–5 demonstrated higher post

Table 5. Pre and post artifact data (Means) by subgroup.

	<i>n</i>	Potential of Task		Implementation of Task*		Teacher Expectations*	
		Pre	Post	Pre	Post	Pre	Post
Overall	50	2.60	2.86	2.20	2.58	2.26	2.65
Grade K-5	30	2.63	2.80	2.23	2.60	2.36	2.71
Grade 6–12	20	2.55	2.95	2.15	2.55	2.11	2.56
Practicum	10	2.70	3.20	2.10	2.80	2.30	2.80
MLC	19	2.42	2.95	2.05	2.63	2.22	2.72
MLT	31	2.71	2.81	2.29	2.55	2.29	2.61

Only matched pre and post data are included. $n = 45$ – 50 for all indicators. Possible Ratings: Potential of the Task 0–4, Implementation of the Task 1–4, and Teacher Expectations 0–4 or not applicable. Wilcoxon tests were used to compare pre to post for overall scales only, not subgroups. *Differences were deemed statistically significant if $p < 0.05$.

scores for two of the three components compared to those teaching in Grades 6–12, and for all three components the mathematics leadership practicum subgroup had the highest post score.

By the end of the project, observation data indicated EaMML teachers were using more research-proven practices in their classroom than they were at the onset of the project. Teachers ($n = 26$) were observed at the beginning and end of the project and were rated in ten areas: observation scores increased over time in all ten areas and significantly in three areas: Questioning, Student Linking, and Providing (Students' Responses). Teachers' scores were highest for Participation and Potential of the Task with mean post scores of 3 or higher as shown in Table 6. Teachers' scores were lowest for Student Linking; the mean was 1.77 at post.

The research team examined key subgroups to determine if there were differences. An examination of subgroup difference reveals that the MLC post scores were higher than the MLTeam post scores for six of the ten components, those teaching Grades K–5 had higher post scores for eight of the ten components compared to those teaching Grades 6–12, and those who completed the mathematics leadership practicum did not have higher scores than other groups for any components.

Research Question 4: Student Learning

Design

Research Question 4 was addressed using a quasi-experimental study. The initial sample was composed of 47,672 records representing 21,496 students in Grades K through 11, in two participating school districts. A total of 692 duplicate records were removed from the sample.¹ Student membership in the treatment or comparison group was determined based on receiving instruction from an EaMML teacher. In the participating districts, a total of 56 teachers served as EaMML teachers ($n_{\text{District 1}} = 20$, $n_{\text{District 2}} = 36$) and 328 ($n_{\text{District 1}} = 148$, $n_{\text{District 2}} = 180$) served as comparison teachers. EaMML teachers engaged in project professional development during the final four months of the baseline year (2014–2015). As such, students taught by an EaMML teacher during this time period ($n = 3,579$) were removed from the analytic sample to ensure that students included in the treatment group were exposed to at least one full year of instruction from an EaMML teacher. Students were included in the treatment group if they were taught by an EaMML teacher in at least one of the treatment years (2015–2016 or 2016–2017) and students were included in the comparison group if they did not receive any instruction from an EaMML teacher during this time. The final analytic sample included data from the students of 287 teachers, 36 EaMML teachers ($n_{\text{District 1}} = 14$, $n_{\text{District 2}} = 22$), and 251 comparison teachers ($n_{\text{District 1}} = 94$, $n_{\text{District 2}} = 157$).

Analysis

Two HLM Models were used to assess the impact of EaMML on student achievement as measured by the Smarter Balanced Assessment of Mathematics. HLM Model 1 was a two-level longitudinal model (i.e., growth model with scores nested within students) that tested whether student achievement scores increased at a greater rate if taught by an EaMML teacher versus a comparison teacher and included a time by treatment interaction term. HLM Model 2 included 3-way interaction to test whether the EaMML effect was moderated by student subgroups. All models included grade level of the student in 2015, district, and controlled for baseline achievement.

¹All duplicate records included identical Smarter Balanced Assessment scores. In 356 of these records a student was linked to multiple teacher types. In these cases, if a student was assigned to an EaMML teacher the record for the EaMML teacher was maintained in the sample. For those with both dropped teachers and non-EaMML teachers the record for the dropped teacher was maintained.

Table 6. Pre and post observation data (Means) by subgroup.

	n	Potential of Task		Implementa- tion of Task		Student Discussion Following Task		Questioning*		Mathematical Residue		Participation		Teachers Linking		Students Linking*		Asking Teachers Press		Providing Students Responses*	
		Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Overall	26	2.81	3.08	2.46	2.65	1.96	2.23	2.27	2.88	1.85	2.15	2.92	3.16	2.12	2.23	1.38	1.77	2.35	2.62	2.08	2.54
Grade K-5	8	2.75	3.13	2.38	2.75	2.25	2.63	2.50	3.00	2.13	2.50	2.75	3.25	2.13	2.38	1.50	1.75	2.75	2.75	2.13	2.38
Grade 6-12	12	2.85	3.00	2.62	2.69	1.77	1.85	2.15	2.77	1.69	1.92	2.75	3.08	2.00	2.15	1.38	1.85	2.15	2.62	2.00	2.62
Practicum	5	3.00	3.00	2.60	2.60	2.60	2.60	3.20	3.00	2.40	2.40	2.60	3.20	2.20	2.40	1.80	1.80	3.40	2.60	2.60	2.40
MLC	10	2.90	3.30	2.50	3.00	2.10	2.10	2.40	2.90	2.30	2.30	2.70	3.00	2.40	2.40	1.60	1.80	2.50	2.70	2.10	2.60
MLT	15	2.75	2.94	2.44	2.44	1.88	2.31	2.19	2.88	1.56	2.19	3.70	3.27	1.94	2.13	1.25	1.75	2.25	2.56	2.06	2.50

Only matched pre and post data are included. Possible Ratings 0–4: Potential of the Task, Implementation of the Task, Student Discussion Following the Task, and Mathematical Residue. Possible Ratings 0–4 or not applicable: Questioning, Participation, Teacher Linking, Asking, and Providing. Wilcoxon tests were used to compare pre to post at the aggregate level. The participation n is 25 rather than 26 due to a score of *not applicable* on the pre. *Differences were deemed statistically significant if $p < 0.05$.

Results

The results of the first HLM model are presented in Table 7. Smarter Balanced Assessment scores for all students increased significantly over time with, on average, scores increasing 31.4 points at each subsequent administration. In addition, students who had an EaMML teacher at any point during the study, scored significantly higher than those of comparison teachers with EaMML students scoring, on average, 23.6 points higher at each administration (Hedges $g = .02$). A significant time by EaMML interaction was also observed with scores increasing, on average, 7.72 points more per administration for EaMML students than for comparison students. The covariate for district was also significant with students in District 1 scoring, on average, 11.24 points higher than students in District 2. The grade level in 2015 indicators were also significant predictors in the model. This is to be expected as scores on the Smarter Balanced Assessment are scaled to increase with each grade level. The predicted scores at baseline, first, and second follow-up for EaMML and non-EaMML students by grade level and district are presented in Table 7 for students in Grade 3, Grade 4, Grade 5, and Grade 6 at baseline.

The second HLM model tested whether the EaMML effect differed by student subgroup, results for this model are presented in Appendix A. The EaMML effect was consistent across students by gender, race (Underrepresented vs. White, Asian, Pacific Islander), ethnicity (Hispanic vs. Non-Hispanic), and SES (Free-and-Reduced Lunch vs. Non). After controlling for demographic variables a significant effect of having been instructed by an EaMML teacher was observed with EaMML students scoring, on average, 22.91 points higher than non-EaMML students. The time by EaMML interaction was also still significant with scores increasing, on average, 6.13 points more per administration for EaMML students than for comparison students. District and the grade level indicators were also still significant in this model.

Among the 2,515 students who did not pass their baseline Smarter Balanced Assessment, students taught by an EaMML teacher ($n = 832$) had greater odds (0.44) of passing a follow up Smarter Balanced Assessment than comparison students ($n = 1,683$; 0.27). The odds ratio (2.20) indicates that students not passing their baseline Smarter Balanced Assessment and were taught by EaMML teachers in subsequent years were over 2 times more likely to pass a future Smarter Balanced Assessment than their peers who were not taught by EaMML teachers in subsequent years.

Study Limitations

There are limitations in this study that could be addressed by future research. Limitations of the teacher-level data include: the teachers were volunteers, there is not a comparison group for teacher-level data, the overall sample size is small, and data were collected pre/post rather than continuously over the course of the project. The survey was self-report data and participants may have reported socially desirable answers. In terms of assessing student achievement, the study was limited to the

Table 7. EaMML effect model.

	Estimate	Standard Error			
	Fixed Effects		<i>t</i>	<i>df</i>	<i>p</i>
Intercept	2341.64	8.28	282.69	4120.70	***
EaMML	23.60	3.15	7.49	3848.80	***
District	11.24	2.95	3.81	3708.86	***
Grade 3	55.40	8.46	6.55	4024.10	***
Grade 4	85.79	8.55	10.04	4019.69	***
Grade 5	123.17	8.68	14.16	4011.93	***
Grade 6	146.38	8.87	16.49	3999.72	***
Grade 7	203.44	9.19	22.14	4006.77	***
Time	31.41	0.67	47.02	3768.80	***
Time * EaMML	7.73	1.05	7.34	3551.22	***

*** $p < .001$.

grades in which statewide assessments were administered so the student achievement data was not representative of all the teachers in the project.

Discussion

There are several unique features of the EaMML project that add to the research base. One unique feature of the EaMML project is the K-12 representation of math specialists engaging in both across-the-grades and within-grade-band professional learning. The EaMML project advances the field by providing an example of a professional development design that strengthens MTL content knowledge, pedagogical knowledge, and leadership knowledge and skills while engaging leaders throughout an entire district and across all grade levels. Yet questions remain about effectiveness of the model at scale that is, in varied contexts, with different facilitators (Borko, 2004; Marrongelle et al., 2013), and with earlier expectations for teacher leadership (c.f. Implementing the Problem-Solving Cycle (iPSC), Koellner & Jacobs, 2015). It would be worthwhile to examine components of the EaMML professional development model through the lens of the professional development continuum described by Koellner and Jacobs (2015). When considering this continuum from “highly specified” to “highly adaptive,” the overall model might fall somewhere in the middle because there is a mix across the PD components in terms of level of specification. As an example, the MS courses are more highly specified than the lesson studies which may have common broad professional learning goals but varied greatly by grade band, task selected, classroom learning environment, and student discourse levels. An aspect of the mix of professional learning experiences may contribute to EaMML’s success. Future research could study various professional development models to determine if this model, or others, are more effective in terms of developing and sustaining leadership structures over time.

The second unique feature of this project is studying the preparation and positioning of classroom teachers as mathematics leaders, as most of the research focuses on mathematics specialists as coaches. In this study, classroom teachers engaged in informal (i.e., sharing resources with colleagues) and formal (i.e., leading professional development sessions) leadership efforts within their school, district, and outside of their district. As noted earlier, coaches are an added expense for schools and often remove the highly skilled teachers from their work with students, so the findings from this research are critical to show that these leadership roles can be effectively filled by classroom teachers and under what conditions. The responses from MLC MLTs two years after completion of the project offer preliminary evidence for how schools or districts might successfully support and sustain such engagement. Future research could study classroom teachers as leaders and address some of the limitations of this study by including a comparison group for teacher-level data, collecting data more frequently than pre and post, and collecting instructional practice data from teacher leaders’ colleagues.

Few studies examine the impacts of math teacher leaders on student achievement. Prior research typically focused on school level data or included student level data with several limitations and primarily examined the role of leaders in a coaching role, not as a classroom teacher. This research examined classroom teachers as leaders, utilized student-level state assessment data in a rigorous quasi-experimental design, included data across elementary and middle school (grades 3–7), and incorporated an equity perspective in the analysis. Additionally, as called for by Sloane and Wilkins (2017), this research employed a quantitative methodology and hierarchical linear modeling (HLM) to assess student achievement gains over time. Students with an EaMML teacher scored significantly higher on the state assessments than students of comparison teachers and the EaMML effect was consistent across students grouped by gender, race, ethnicity, and SES. By using HLM, this research was able to make a unique contribution to the field by connecting leadership to student achievement. Future research should utilize these types of rigorous methods to analyze impacts on student achievement and should ensure models are included that address potential differences in achievement among subgroups (e.g., gender, race, ethnicity, and SES). Use of these models are necessary to determine if an intervention was able to minimize or close achievement

gaps. Due to the limitations of state-standardized assessments, future research could use common school or district assessments that may be more closely aligned with the project's student learning goals.

What is less clear is the relationship between student achievement and research-proven practices. Project leaders defined strong mathematical practices as scores of 3 and 4 on the IQA as descriptions at these levels are consistent with the research-proven effective mathematics teaching practices (NCTM, 2014) and equity-based practices (Aguirre et al., 2013) as they go deep with mathematics (Academic Rigor rubrics), assigning competence as well as affirming and positioning students as doers, knowers, and sense makers (Accountable Talk rubrics). The artifact and observation data reveal that even though average scores were not well within the range of strong mathematical practices at the time of the post assessment, there was a significant positive impact on student achievement. This finding raises questions regarding the level, frequency, and duration of strong mathematical practice needed to impact students' opportunities to learn. It also raises questions about the content, structure, frequency, and duration of the professional learning that would be needed for more post scores to be in the strong mathematical practices range. Furthermore, might this be related to a disconnect between expectations of students on the Smarter Balanced Assessment and strong mathematical practices with the latter pressing further with regard to rigor?

Finally, this research was unique because it collected artifact and observation data while most of the previous research has focused on observations. Consistent with many studies of classroom practice, the data reveal stronger instructional practice with task potential than with task implementation (M. D. Boston & Smith, 2009; Henningsen & Stein, 1997; Stein & Lane, 1996) for both measures, however stronger performance was evident in the observation. This difference observed between the artifact and observation was not expected, in that observations represented one point in time of a lesson, artifacts could be collected from many lessons potentially leading to stronger performance because the teacher had more control over what they submitted. Project leaders speculated the weaker performance on the post artifact, when compared to the observation, could be related to some participants not taking the final data point as seriously as the other data points. The difference could also be attributed to the fact that the observations were from only a subset of teachers, so even though the observations were of randomly selected teachers, this group may not have been representative of the whole. This raises a question about the extent to which the artifact scores can be a predictor of the observation scores for the Potential of the Task and the Implementation of the Task. Project leaders also wondered if there are additional areas explicitly connected to the Accountability to Knowledge and Rigorous Thinking Rubrics (i.e., Asking (Teacher's Press), Providing (Students' Responses)) that could be scored for artifacts (e.g., if student samples included teacher feedback and students' responses), or if the observations are simply a more complete picture of teachers' classroom practice. Conducting observations and engaging in consensus conversations after scoring was more time-consuming for researchers than engaging in similar conversations about classroom artifacts, yet in a large project with fewer resources, collecting more robust classroom artifacts may be a good option.

Conclusion

The EaMML project, in an effort to build district-wide mathematics leadership, engaged both formal (i.e., coaches and administrators) and informal leaders (i.e., elementary teachers teaching all subjects as well as middle- and high-school mathematics teachers) in common professional learning. The EaMML professional learning increased MTLs' mathematical content knowledge, pedagogical knowledge, leadership knowledge, and use of research-based practices. This research showed that students with an EaMML teacher scored significantly higher on the state assessments than students of comparison teachers and the EaMML effect was consistent across students by gender, race, ethnicity, and SES.

Acknowledgments

The authors would also like to thank Jennifer Weston-Sementelli for conducting a thorough statistical review and Caroline Qureshi for her instrumental contributions both during the study and for this manuscript.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Mathematics and Science Partnerships grant program, a competitive grant through the Oregon Department of Education [Office of Elementary and Secondary Education, Title IIB].

ORCID

Nicole Rigelman  <http://orcid.org/0000-0003-1868-3076>

References

- Aguirre, J. M., Mayfield-Ingram, K., & Martin, D. B. (2013). *The impact of identity in K-8 mathematics: Rethinking equity-based practices*. National Council of Teachers of Mathematics.
- Association of Mathematics Teacher Educators. (2013). *Standards for elementary mathematics specialists: A reference for teacher credentialing and degree programs*.
- Balfanz, R., MacIver, D. J., & Byrnes, V. (2006). The implementation and impact of evidence-based reforms in high-poverty middle schools: A multi-site, multi-year study. *Journal for Research in Mathematics Education*, 37(1), 33–64. <https://doi.org/10.2307/30035051>
- Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In G. Sykes & L. Darling-Hammond (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3–32). Jossey Bass.
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, 33(8), 3–15. <https://doi.org/10.3102/0013189X033008003>
- Boston, M. D., & Smith, M. S. (2009). Transforming secondary mathematics teaching: Increasing the cognitive demands of instructional tasks used in teachers' classrooms. *Journal for Research in Mathematics Education*, 40(2), 119–156. <https://doi.org/10.2307/40539329>
- Boston, M., & Wolf, M. K. (2004, April). *Using the Instructional Quality Assessment (IQA) toolkit to assess academic rigor in mathematics lessons and assignments*. Presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.
- Brosnan, P., & Erchick, D. (2010). Mathematics coaching and its impact on student achievement. In P. Brosnan, D. B. Erchick, & L. Flevares (Eds.), *Proceedings of the 32nd annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. VI, pp. 1362–1370). Columbus, OH: The Ohio State University.
- Campbell, P. F., Griffin, M. J., & Malkus, N. N. (2017). Factors influencing elementary mathematics specialists' impact on student achievement. In M. B. McGatha & N. R. Rigelman (Eds.), *Elementary mathematics specialists: Developing, refining, and examining programs that support mathematics teaching and learning* (pp. 193–202). Information Age Publishing.
- Campbell, P. F., & Malkus, N. N. (2013). The mathematical knowledge and beliefs of elementary mathematics specialist-coaches. *ZDM the International Journal on Mathematics Education*. <https://doi.org/10.1086/657654>
- Coniam, S. (2010). Mathematics coaching and its impact on urban fourth grade students' mathematics proficiency on high stakes testing. In P. Brosnan, D. B. Erchick, & L. Flevares (Eds.), *Proceedings of the 32nd annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. VI, pp. 1379–1386). The Ohio State University.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181–199. <https://doi.org/10.3102/0013189X08331140>
- Doerr, H. M., Goldsmith, L. T., & Lewis, C. C. (2010). *Mathematics professional development: Professional development research brief*. National Council of Teachers of Mathematics.

- 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. <https://doi.org/10.3102/00028312038004915>
- Harris, A. (2003). Teacher leadership as distributed leadership: Heresy, fantasy or possibility? *School Leadership & Management*, 23(3), 313–324. <https://doi.org/10.1080/1363243032000112801>
- 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. <https://doi.org/10.5951/jresmetheduc.28.5.0524>
- Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal*, 105(1), 11–30. <https://doi.org/10.1086/428763>
- Hurd, J., & Lewis, C. (2011). *Lesson study step-by-step: How teacher learning communities improve instruction*. Heinemann.
- Koellner, K., & Jacobs, J. (2015). Distinguishing models of professional development: The case of an adaptive model's impact on teachers' knowledge, instruction, and student achievement. *Journal of Teacher Education*, 66(1), 51–67. <https://doi.org/10.1177/0022487114549599>
- Marrongelle, K., Sztajn, P., & Smith, M. (2013). Scaling up pro-fessional development in an era of common state standards. *Journal of Teacher Education*, 64(3), 202–211. <https://doi.org/10.1177/0022487112473838>
- McGatha, M., Davis, R., & Stokes, A. (2015). *The impact of mathematics coaching on teachers and students* (M. Fish, Ed.). NCTM Research Brief. National Council of Teachers of Mathematics.
- Meyers, H. W., & Harris, D. (2008). The Vermont Mathematics Initiative: Student Achievement from Grade 4 to Grade 10, 2000–2006. *James M. Jeffords Center for Policy Research*, <https://scholarworks.uvm.edu/jmjcpr/33>
- Muijs, D., & Harris, A. (2003). Teacher leadership—Improvement through empowerment? *Educational Management & Administration*, 31(4), 437–448. <https://doi.org/10.1177/0263211X030314007>
- National Council of Teachers of Mathematics. (2014). *Principles to actions: Ensuring mathematical success for all*.
- Nickerson, S. D. (2010). Preparing experienced elementary teachers as mathematics specialists. *Investigations in Mathematics Learning*, 2(2), 51–68. <https://doi.org/10.1080/24727466.2010.11790294>
- Sloane, F. C., & Wilkins, J. L. M. (2017). Aligning statistical modeling with theories of learning in mathematics education research. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 183–207). National Council of Teachers of Mathematics.
- Smith, M. S. (2001). *Practice-based professional development for teachers of mathematics*. National Council of Teachers of Mathematics.
- Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation*, 2(1), 50–80. <https://doi.org/10.1080/1380361960020103>
- Sutton, J. T., Burroughs, E. A., & Yopp, D. A. (2011). Coaching knowledge: Domains and definitions. *Journal of Mathematics Education Leadership*, 13(2), 12–20.
- Sztajn, P., Borko, H., & Smith, T. (2017). Research on mathematics professional development. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 793–823). National Council of Teachers of Mathematics.
- Wasley, P. A. (1991). *Teachers who lead: The rhetoric of reform and the realities of practice*. Teachers College Press.
- Watanabe, T. (2002). Learning from Japanese lesson study. *Educational Leadership*, 59(6), 36–39. <https://www.ascd.org/el/articles/learning-from-japanese-lesson-study>
- Weaver, D., Lewis, C., Qureshi, C., Hiebert Larson, J., Gray, M., Wang, X., & Wadeson, K. (2018). *Arizona mathematics project—adding it up: A 5 Year comprehensive report*. RMC Research Corporation.
- Zollinger, S., Brosnan, P., Erchick, D. B., & Bao, L. (2010). Mathematics coaching: Impact on student proficiency levels after one year of participation. In P. Brosnan, D. B. Erchick, & L. Flevares (Eds.), *Proceedings of the 32nd annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. VI, pp. 1379–1386). The Ohio State University.

Appendix A. EaMML Effect Equity Model

	Estimate	Standard Error			
	Fixed Effects		t	df	p
Intercept	2360.24	8.51	277.22	4292.46	***
EaMML	22.91	5.39	4.25	4687.05	***
District	13.54	2.87	4.72	3742.02	*
Grade 3	52.17	8.19	6.37	3989.76	***
Grade 4	83.12	8.28	10.04	3986.01	***
Grade 5	120.30	8.40	14.32	3978.38	***
Grade 6	142.40	8.60	16.56	3968.37	***
Grade 7	199.17	8.90	22.38	3979.33	***
Female	0.35	3.59	0.10	3990.91	n.s.
Underrepresented	−28.13	5.00	−5.63	4300.72	***
Hispanic	−3.03	5.59	−0.54	4200.38	n.s.
FRL Eligible	−5.39	2.32	−2.32	4823.60	**
Time	37.71	1.53	24.69	4068.52	***
Time * Female	1.39	1.33	1.04	3735.42	n.s.
Time * Underrepresented	−5.46	1.93	−2.83	3911.30	***
Time * Hispanic	2.34	2.13	1.10	3891.16	n.s.
Time * FRL	−6.75	1.63	−4.14	4429.83	***
Time * EaMML	6.13	2.32	2.64	3843.62	***
EaMML * Female	2.32	5.86	0.40	3893.24	n.s.
EaMML * Underrepresented	7.40	8.46	0.87	4147.54	n.s.
EaMML * Hispanic	−14.66	9.61	−1.52	4028.38	n.s.
EaMML * FRL	−2.99	3.89	−0.77	4942.26	n.s.
Time * EaMML * Female	1.71	2.11	0.81	3539.30	n.s.
Time * EaMML * Underrepresented	3.61	3.17	1.14	3701.32	n.s.
Time * EaMML * Hispanic	−5.62	3.54	−1.59	3654.89	n.s.
Time * EaMML * FRL	0.11	2.50	0.04	4149.50	n.s.

n.s. non-significant, * $p < .05$, ** $p < .01$, *** $p < .001$.