2021

An Evaluation of the Teaching English Learners Early Mathematics (TEEM) Program; An Investing in Innovations (i3) Development Grant (U411C140073)

IMPLEMENTATION YEARS 1-4 FINAL REPORT

PREPARED BY: DR. LORIE A. SOUSA WITH DR. MADELEINE JETTER (CSUSB)

COREducational

Table of Contents

Executive Summary

California State University, San Bernardino (CSUSB) and the Riverside County Office of Education (RCOE) were awarded an Investing in Innovation (i3) Development grant in 2015 to investigate the impact of an early education mathematics program on English Learner students at two rural school districts in Riverside, CA.

COREducational, an evaluation and research consultancy, conducted the external evaluation of the findings over a four-year period from 2015-2016 to 2018-2019 school years. A no cost extension was approved for the final and fifth year (2019-2020). However, the COVID pandemic resulted in school closures that precluded data collection. Thus, the decision was made to report on the four complete years of data collected.

Project Activities and their relationship to project goals:

Teaching English Learners Early Mathematics (TEEM) had five major goals: to improve the academic achievement of English Learners (ELs) (Goal 1); to improve teacher knowledge and beliefs for teaching math to ELs (Goal 2); to improve the teaching of math to ELs and all students (Goal 3); to develop principals as instructional leaders in mathematics (Goal 4); and to build a learning partnership that develops, sustains and shares a model of instructional change supporting ELs. (Goal 5). In support of these goals, a combination of four mutually supportive interventions were implemented.

Intervention 1. Intensive Teacher Professional Development (PD) Activities

Summer Intensive Institutes were held every summer. The institutes addressed math content, best practices for teaching English Learners, and interactive notebooking. Over the course of the project, the mathematical focus of the Institutes progressed through the major themes of elementary mathematics, including whole numbers and algebraic thinking, multiplicative thinking, geometry, fractions, proportional reasoning and expressions and equations. The Common Core State Standards for Mathematical Practice were emphasized throughout the professional development. The use of multiple representations, to support understanding for English Learners and all students, was another focus of attention. Participants received curriculum units from the Contexts for Learning Mathematics curriculum as a resource to support them in teaching through problem solving. TEEM staff, including CSUSB faculty and Riverside County Office of Education (RCOE) specialists, presented the PD. Lesson Study experts were also brought in to cover special topics such as Teaching Through Problem Solving.

Intervention 2. Lesson Study

Each year, participants worked in grade-level groups of three to eight members each to complete two three-day cycles of lesson study in each academic year of the project. In a lesson study cycle, teams set shared instructional goals; studied relevant content, pedagogy and formative assessment data; and planned, taught, observed and reflected on a mathematics lesson. As a result of participation in the California Action Network for Mathematics Excellence and Equity (CANMEE), lesson study participants began crafting equity goals (in addition to mathematical goals) for the

classrooms in which lesson study is taking place. An example of an equity goal is "Students see themselves as powerful mathematical thinkers through small-group and whole-classroom discourse."

Intervention 3. Interactive Math Notebooks for participating students

Each participating classroom teacher received a class set of composition books at each TEEM Summer Institute. Notebooks are intended to contain a record of the teacher's teaching and the student's learning: "input" pages on the right hand side include note-taking, practice, or other elements directed by the teacher, while "output" pages on the left provide opportunities for students to engage more actively with the material through problem solving, summarizing or reflection on learning. Project staff rated a sample of student notebooks each year using a rubric.

Intervention 4. Professional Learning for school leaders

Site principals participated in several activities designed to support them in understanding the curriculum shifts described in the Common Core State Standards for Mathematics.

TEEM staff provided 30 to 45-minute presentations at monthly district-based principals' meetings (six at each district). These presentations addressed the practices and strategies implemented in TEEM to aid principals in understanding what they would expect to see in observations of TEEM classrooms. Topics included Lesson Study, mathematics notebooking, and the importance of multiple representations in mathematics learning, among others.

All principals participated in classroom walkthroughs with TEEM staff. In one of the districts, each site principal hosted one walkthrough attended by all district principals and TEEM staff. In the other district, the two principals walked classrooms at both sites twice a year. The *5x8 Card* developed by the Strategic Education Research Partnership (SERP) was used to facilitate observation and discussion of mathematics classroom practice. This led to more productive conversations with principals about mathematics teaching and learning.

Independent Evaluation and preliminary findings:

The Independent Evaluation of the TEEM Project was conducted by TEEM's independent evaluator COREducational, headed by Dr. Lorie Sousa. Below is a summary of results.

TEEM Teachers demonstrated growth in mathematics content knowledge, as measured in the Learning Mathematics for Teaching (LMT) Assessment. They have showed improving self-efficacy beliefs relative to a comparison group.

Participating teachers' mathematics content knowledge improved relative to a comparison group of non-participating teachers in each project year (although due to sample size, the difference was not statistically significant in the last two project years).

The Mathematics Teaching Efficacy and Expectancy Beliefs Instrument (MTEEBI) was used to measure teachers' beliefs about developing mathematical practices in a diverse student population. Teachers are asked to rate their level of agreement with statements like "I am able to help students from impoverished backgrounds excel in math." Participating teachers' beliefs improved relative to a comparison group to a statistically significant degree, showing that participation in TEEM had a measurable, positive impact on teachers' beliefs -- both in their own teaching abilities and in their students' ability to learn mathematics.

Classroom observations were performed each project year using a project-specific rubric. These results were analyzed in the evaluation report and demonstrated some significant positive shifts in teacher classroom practices. For example, the percentage of classrooms in which "English Learners produce language that communicates ideas and reasoning, even when that language is imperfect" increased from 48% in Spring 2016 to 80% in Spring 2018.

Student outcomes were tracked using state-administered Smarter Balanced Assessment (SBA) for grades 3 and higher. Because students do not take these tests prior to grade 3, we used Assessing Math Concepts (AMC) interviews for grades PK-2. In grades 3 and above, students taught by a TEEM teacher for at least one year outperformed students in a comparison group in both the Mathematics and English Language Arts SBA tests. This was true both for English Learners and the general student population. Comparison data was not available for AMC interviews, so our evaluator compared participating students' results at the beginning and end of each project year and found significant improvements in every section of the assessment.

Extent to which objectives, expected outcomes and annual targets were achieved:

Based upon the evaluation data, TEEM achieved many of its proposed outcomes for teachers and students. Due to participant attrition, participation levels were below the proposed target of 75 teachers participating for all four years. Some objectives for site principals could not be achieved as proposed: for example, principal attendance at the TEEM Summer Institute remained well below target in spite of an increase in Y4.

Sustainability planning and dissemination:

The TEEM Teacher Leadership Team consisted of eight teachers interested in developing themselves as math leaders and disseminating best practices learned in TEEM. Four bimonthly Leadership Team seminars were held in Y4. The Project director lead each meeting using the *Learning to Lead Mathematics Professional Development* curriculum by Carroll and Mumme, and participant feedback was very positive. Based on teacher interest, Leadership Team teachers provided activities and material kits to support Family Math nights at school sites. The leadership team also went on to plan and present an online dissemination workshop attended by approximately 30 local PK-TK-Kindergarten teachers at the close of the project.

Project dissemination goals were met. The Project Director gave two presentations at regional conferences discussing the math notebooking practice in TEEM classrooms. The first of these was done in collaboration with a project teacher at the Riverside County MaTHinK conference; the other was to the California Math Council Southern Section Conference. An article submission to NCTM on the same topic was not accepted for publication. Alternate publication venues are being considered to disseminate information regarding the notebooking practice and results of the study.

Challenges, best practices and lessons learned:

Attrition. Teacher attrition was a challenge. In the hope of compensating for attrition, TEEM received permission to invite new teachers to participate in lesson study beginning in Fall of 2018. As a result,10 new teachers were included in TEEM lesson study in September 2018.

Substitute teacher availability. Occasionally, a school district partner was short on substitute teacher coverage on a TEEM lesson study day. In these situations, administrators did their best to enable teachers to keep their lesson study meeting (for example, by splitting classes into other teachers' classrooms), however on some occasions district administrators cancelled a day of TEEM lesson study. This was difficult for teachers and facilitators alike. TEEM attempted to mitigate this problem by scheduling lesson study well in advance.

Lesson study. Work with the CANMEE Lesson Study Support Hub provided insight into best practices for mathematics lesson study. The use of equity goals and an equity commentator also encouraged a deeper look into student learning, causing teachers to slow down and think specifically about targeted students including English learners or those with special needs.

Impact beyond the partners and implementation sites:

Lesson study. Through the CANMEE Lesson Study Support Hub, TEEM staff support lesson study across our region as a vehicle to promote equitable and rigorous mathematics instruction. Further, educators involved in the project have developed their own skills and capacity with respect to the lesson study practice.

Introduction

The *Teaching English Learners Early Mathematics (TEEM)* program was designed by, California State University, San Bernardino (CSUSB) and the Riverside County Office of Education (RCOE), in conjunction with two Local Education Agency (LEA) partners, and the Head Start and State Preschool programs affiliated with one of the LEAs. The purpose was to build on a history of close collaboration to develop, implement and evaluate a replicable model for improving the academic achievement of English Learner (EL) students in pre-K through grade 6.

TEEM program intentions were to increase the number and proportion of ELs completing challenging courses in core academic subjects through a novel combination of interventions designed to support communities of inquiry among students, teachers, and school leaders. Students engaged in structured inquiry through the use of interactive notebooks in mathematics; teachers sought to improve mathematics instruction through the formation of communities of inquiry supported by Summer Institutes and an adaptation of Japanese Lesson Study; and principals were developed as instructional leaders in mathematics through a variety of supports. The project intended to directly impact at least 4,000 students, 75 teachers and 7 principals.

The primary goal was to increase the number and proportion of ELs successfully completing challenging courses in core academic subjects by implementing structured inquiry-based approaches to learning, teaching and school leadership. These approaches were intended to develop collective efficacy in schools and provide ELs access to challenging math content from pre-K through grade 6.

Importance of math in early grades.

To maximize long-term impact, math-focused interventions were developed for ELs in early grades. The risk of dropout is roughly twice as high for ELs as it is for non-ELs (Callahan, 2013), and in many cases the dropout problem is a math problem. In the Los Angeles Unified School District, students who passed Algebra I by grade 9 were twice as likely to graduate as those who had not (Silver, 2008). In order to improve academic outcomes through high school, ELs must succeed in challenging math courses, and they must begin early. Academic performance in early grades is a strong indicator of future dropout risk, and high-quality preschool decreases the risk (Rumberger & Lim, 2008).

In California, achievement gaps existing in grade 2 largely persist through grade 7, a "pivot year" in which student achievement varies widely and determines future opportunities in math (Terry and Rosin, 2011). In California, ELs are a large percentage of early-grade students, and many struggle in math. In 2013, ELs made up 35.7% of CA grade 2 enrollments, and 47% of those students tested below proficient in math.

The status of ELs statewide is mirrored in the partner, small districts serving the unincorporated communities in Riverside County. The percentage of students receiving free or reduced-price lunch is 76.1% and 65% at the two districts, while the state average is 58%. Per the 2010 Census, the largest ethnic group in both communities is Hispanic/Latino at 51.4% and 54.5%. In one district involved in the study, 45.4% of people aged 5 or more years speak a language other than English at home. Results from the

2013 California Standards Test (CST) showed the percentage of students classified as EL tapers through the elementary grades.

While the proportion of ELs decreases, the math achievement gap for ELs grows with the complexity of the mathematics from grade to grade. In the seventh grade, 52% of English-Fluent or English Only seventh graders were proficient or advanced in mathematics while only 13% of ELs attained that level of achievement. A student classified as EL in the grade 7 pivot year faces restricted access to advanced mathematics, to success in high school and college, and to STEM careers.

Mathematics interventions supporting ELs and their teachers are especially timely in light of the Common Core State Standards for Mathematics (CCSS-M). The CCSS-M call for students to demonstrate proficiency through eight Standards for Mathematical Practice (SMP). In order to give ELs access to the CCSS-M, Moschkovich (2013) recommends that teachers balance concepts with procedures; maintain high cognitive demand; develop productive student beliefs about math; and engage students in the SMP. At the same time, teachers in the 22 states using the new Smarter Balanced Assessment Consortium (SBAC) tests to assess student proficiency with respect to the CCSS-M are challenged to prepare ELs for the higher language demands of these tests. TEEM supported teachers in this shift by developing a replicable model of intensive professional learning.

Improving outcomes in core academic subjects through attention to metacognitive and non-cognitive skills and mindsets.

When students believe that intellectual ability is not fixed but grows through effort, they show greater persistence (Dweck, 2006). The National Mathematics Advisory Panel (2008) reported that interventions addressing affective and motivational factors can improve the achievement of traditionally underserved groups. TEEM will help ELs develop the metacognitive and non-cognitive skills key to success using interactive inquiry notebooks for math. The notebooks were not only a tool for collecting and recording classroom activities, but they sought to make student metacognition explicit and provide a structured space for developing reading, writing, and study skills.

Unlike many accepted practices for supporting ELs which focus on general academic language development, the program sought to achieve improved outcomes for ELs while keeping a strong disciplinary focus on *mathematics*, combining four structured inquiry-focused interventions targeting students, teachers, and school leaders. TEEM's approach to improving the achievement of ELs is based on the novel premise that educators need deep *mathematical* understanding to provide ELs access to the SMP in a CCSS-aligned curriculum, that "vocabulary is necessary, [but] not sufficient" to provide that access (Moschkovich, 2013), and that providing that access will support ELs in strengthening language skills through reasoning, talking, and writing about complex disciplinary content.

To provide a comprehensive approach to increasing ELs' access to rigorous mathematics, the project combined four evidence-based interventions:

1. The adaptation of interactive inquiry notebooks – a best practice in science teaching – to mathematics

2. Summer Professional Development (PD) to immerse teachers in math inquiry and notebooking

3. The adaptation of the Japanese Lesson Study model of professional learning for teachers

4. The purposeful, regular, and intensive participation of school principals in the professional development.

Given the need to intervene early, another novel aspect of TEEM was the creation of a PD pipeline for preschool educators who rarely receive opportunities for math-focused PD. Through its focus on mathematics, its combination of research-based interventions, and the purposeful inclusion of school leaders and early education providers, TEEM intended to provide a blueprint for how to increase EL student academic performance and pursuit of STEM in college and careers. TEEM interventions were built upon existing research and promising practice, with adaptations intended to make it an exemplar for new practice.

Interactive notebooks.

Notebooks have been used successfully in many science classrooms (Hargrove & Nesbit, 2003). Used intentionally, student notebooks provide evidence of the teacher's teaching practices and the students' learning (Shimojyo & Shimojyo, 2013) and are a formative assessment tool of great value. They make student metacognition explicit and provide a structured space for the development of reading, writing, and study skills. It is a promising practice, and an inexpensive and sustainable one since paper notebooks are easily obtained. It is a culturally responsive practice: in Mexican and other Latin American classrooms, *el cuaderno* (notebook) is used in a more structured and rigorous fashion than is common in the United States (Celedon-Pattichis & Ramirez, 2012). Amaral and colleagues (2002) found that the use of interactive notebooks in elementary science classrooms in an impoverished section of inland California improved EL achievement not only in science, but also in reading and writing. TEEM hopes to move the field forward by adapting a promising practice in science education to the mathematics classroom and measuring its impact. We tested the idea that use of purposefully designed math inquiry notebooks would help develop language, organizational, and metacognitive skills, and would accelerate EL achievement not only in math but also in language learning.

The combination of interactive notebooks with the lesson study model for PD is a unique and powerful way for teachers to develop effective formative assessment practices.

Lesson Study.

The transition to CCSS calls for teachers to make major shifts towards greater focus, coherence, and rigor in math instruction (CCSS Initiative, 2014), but observers have noted that fundamental teaching practices have changed little in the last century (National Research Council, 2001). TEEM supported teachers in this shift using the Japanese lesson study model.

Lesson study gained attention with the publication of *The Teaching Gap* (Stigler & Hiebert, 1999) and is the subject of active research (Lewis et. al. 2009). Lesson study incorporates many characteristics of effective PD and school change: it is intensive, ongoing, and connected to practice; it is focused on student learning, it addresses the teaching of specific curriculum content, and it builds strong working relationships among teachers (Darling-Hammond et. al. 2009). In lesson study, teams of teachers develop a practice of inquiring into their students' thinking and investigating the effectiveness of their instruction. In a cycle of lesson study (Fig. 1), teachers agree to shared learning goals. They research relevant curriculum and pedagogy, and they plan, observe, and refine a "research lesson" addressing the goals.

Based on its documented effectiveness and on experience gained in previous projects, TEEM used lesson study to create new opportunities for schools to build a knowledge base for teaching core academic subjects to ELs. Through close observation, teachers develop "the eyes to see students" (Lewis, 2004), seeing past language deficiencies into the content of students' mathematical thinking. TEEM teachers sought to build and share knowledge of their students' language needs and learn to anticipate those needs. Student notebooks were a key resource for lesson study, providing teachers with data to inform their study, help discern student needs, set instructional goals and evaluate success of research lessons.

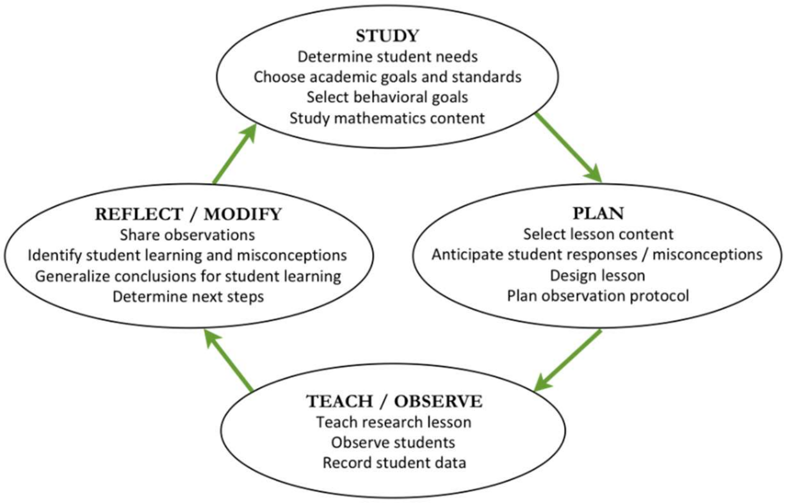

STUDY
Determine student needs
Choose academic goals and standards
Select behavioral goals
Study mathematics content
PLAN
Select lesson content
Anticipate student responses / misconceptions
Design lesson
Plan observation protocol
TEACH / OBSERVE
Teach research lesson
Observe students
Record student data
REFLECT / MODIFY
Share observations
Identify student learning and misconceptions
Generalize conclusions for student learning
Determine next steps

Figure 1 - The Lesson Study Cycle

Supporting Instructional Leadership in Mathematics.

Among all school-related factors, the impact of leadership on student learning is second only to teaching, and this impact is greatest in high-need schools. At the same time, there is little research on how principals continue their professional learning on the job (Leithwood et. al, 2004). While *Instructional Rounds* (City et. al, 2009) are considered best practice for principal learning, principals without a significant background in mathematics receive little support in considering the mathematical content of the lessons they observe in classrooms. To fill this gap, TEEM's fourth novel intervention sought to support principals as "lead learners" at schools (Liptak, 2005), developing their capacity as instructional leaders in the transition to CCSS. Principals attended Summer Institute sessions and monthly sessions built into district-wide principal meetings. Project staff supported principals in Instructional Rounds.

The proposed interventions and the evaluation of their effectiveness will hopefully benefit schools and districts nationwide: EL populations are growing rapidly across the nation (by 53% from '98 to '08; NMAP, 2008) and, to our knowledge, research investigating methods to accelerate EL growth via CCSS, lesson study, and interactive notebooking has not yet been undertaken.

Method

Project Design

Overview and Goals. The TEEM program design is based on Project DELTA (2010-2013 California Math and Science Partnership (CaMSP)), a partnership between RCOE, CSUSB, and five Riverside County LEAs. Project DELTA focused on intensive PD paired with lesson study to improve teacher mathematical knowledge and student achievement. Three years of student achievement data were used to investigate project impact. This pilot study met the ED-MSP's criteria for rigor and showed that intervention students outperformed a comparison group on state achievement tests. Notably, *English Learners* in program classrooms showed even greater growth than students in the general population.

At the close of Project DELTA, an independent evaluator, conducted a summative evaluation. The percentage of students improving their achievement by one performance level or more on the CST (grades 3 – Algebra I) was recorded for each year of the project. The results are shown in Figure 2:

In each year of the project, students of *DELTA* teachers improved at a greater rate than their peers in a Comparison Group, and that rate of improvement was sustained through all three years of the project. Performance data for 2011-2012 was disaggregated by three major subgroups (Students with Disabilities, English Learners, and National School Lunch Program participants (a proxy for low socio-economic status)) as shown in the table to follow. High-need students showed more improvement than the overall student population, and all subgroups outperformed their peers in the Comparison Group.

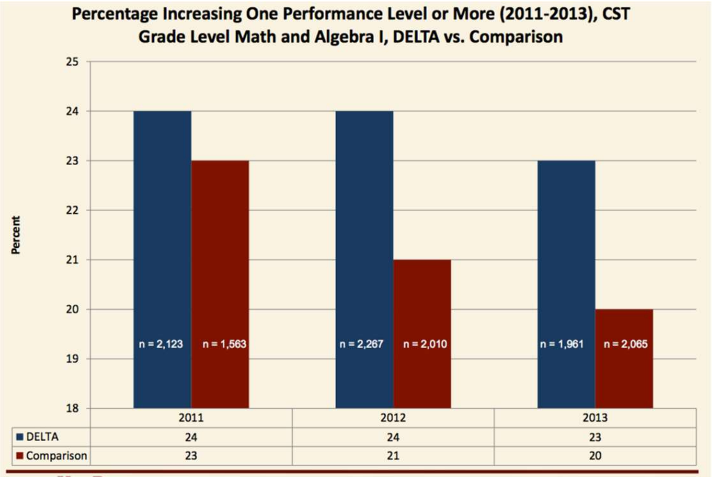


Figure 2 – Project DELTA Results

Table 1: DELTA CST Performance by group

Percentage of Students increasing one CST performance level or more (2011-2012), Grade Level Math and Algebra I, *DELTA* students vs. Comparison				
	Students w/ Disabilities	English Learners	NSLP Participants	All Students
DELTA	32.2	27.9	25.5	24.3
Comparison	29.6	24.4	22.4	20.9

Table 2: Effect Size CST ELA, Math, and CELDT

Effect Size s- Cohen's *d*						
Outcomes	n_I	n_C	*p*-value		Cohen's *d* effect size	Small, Moderate, or Large Effect*
CST ELA PL 2007	435	662	0.02		0.14	small
CST ELA PL 2008	499	584	0.06		0.12	small
CST ELA PL 2009	488	523	.000		0.29	small
CST ELA PL 2010	534	503	.000		0.35	moderate
CST ELA PL 2011	452	395	.000		0.41	moderate
CST Math PL 2007	435	659	.000		0.39	moderate
CST Math PL 2008	498	569	.000		0.31	moderate
CST Math PL 2009	485	508	.000		0.47	moderate
CST Math PL 2010	523	479	.000		0.51	moderate
CST Math PL 2011	448	377	.000		0.42	moderate
CELDT Change from 2006 to 2011	159	115	0.03		0.27	small

CST = California Standards Test; CELDT = California English Language Development Test

As can be seen in Table 2, small to moderate effect sizes were obtained in the final summative analysis of Project DELTA, providing substantive evidence of program efficacy. Expanding upon this work, TEEM set out to strengthen the impact on English language development through the addition of interactive mathematics notebooks in early grades and through the inclusion of preschool educators, thus, it was an extension of the previously implemented project.

In crafting a comprehensive approach to improving outcomes for ELs, an overarching focus on communities of inquiry was maintained for students, teachers, and principals.

Goal 1. Improve the academic achievement of ELs.	
Outcome 1.1	<u>Student mathematics performance</u>: Students will improve their mathematics performance as measured by SBAC assessments (grades 3-6) and Assessing Math Concept (AMC) Interviews (Richardson, 2011) (pre-K - 3).
Outcome 1.2	<u>EL mathematics achievement</u>: EL students will reduce the mathematics achievement gap relative to their non-EL peers, as measured by SBAC assessments (grades 3-6) and AMC Interviews (pre-K - 3).

Outcome 1.3	Student language and metacognitive skills: ELs will show greater growth in reading and writing skills and in habits of mind compared to students whose first language is English, as measured by SBAC-ELA.
Outcome 1.4	Student course-taking patterns: EL students in Middle School will complete and succeed in advanced math courses at higher rates, as seen in enrollment data.
Goal 2. Improve teacher knowledge and beliefs for teaching math to ELs.	
Outcome 2.1	Teacher content knowledge: Teachers will deepen their mathematical knowledge for teaching as measured by the LMT instrument.
Outcome 2.2	Teacher beliefs: Teachers will improve their self-efficacy for teaching the CCSS-M, as measured by the MTEEBI instrument (Riggs et. al., 2018)
Goal 3. Improve the teaching of math to ELs and all students.	
Outcome 3.1	Cognitive Demand: Math lessons will maintain higher levels of cognitive demand and align with CCSS-M content and practices, as measured in classroom observations.
Outcome 3.2	Teacher Formative Assessment: The frequency and quality of teachers' formative assessment will improve, as measured through observations, surveys and checking student notebooks against a project-developed rubric.
Goal 4. Develop principals as instructional leaders in mathematics.	
Outcome 4.1	Principal content knowledge: Principals will deepen their mathematical knowledge for teaching as measured by the LMT instrument.
Outcome 4.2	Principal instructional leadership: Principals will demonstrate growth as shown on a modified Principal Instructional Management Rating Scale (Hallinger & Murphy, 1985)

MTEEBI = Mathematics Teaching Expectancy and Efficacy Beliefs Instrument

The logic model in the appendix illustrates TEEM's theory of action.

Activities. The following activities were implemented over four years of the program:

Activity 1. Interactive math inquiry notebooks for students (supporting Goal 1). Through Summer PD and Lesson Study, TEEM supported teachers and, in turn, their students, in implementing interactive inquiry notebooks for math. Every student whose teacher participated in TEEM interventions was expected to maintain a hardbound notebook and use it on a daily basis as a personal reference on mathematics vocabulary, to organize their mathematics work, and to keep a summative record of their thinking and problem solving in CCSS-aligned mathematics. Notebook implementation began in August 2015 for the 2015-16 school year. TEEM implemented developmentally appropriate adaptations of the interactive notebook in preschool classrooms following the "authentic assessment" model (Epstein et. al., 2004).

Activity 2. Intensive summer professional development (Goals 2 & 3). To teach ELs equitably, teachers must develop a body of specifically *mathematical* knowledge. For example, a teacher with a deep understanding of place value and properties of arithmetic knows that the US standard algorithm is just one of many correct ways to multiply two-digit numbers, and that students could come into the classroom using any of them.

The mathematical content of the Summer institutes was based on the sequence of key ideas described in the CCSS-M learning progressions documents for the elementary grade levels. The program devoted special attention to promoting robust understanding of novel and challenging topics in the new standards, including counting and cardinality; early algebraic thinking embedded in whole number operations based on ideas of Cognitively Guided Instruction (Carpenter et. al., 1999); Fractions on the number line; and modeling with mathematics.

Teaching practices facilitating EL access to the SMP were built into institute activities, including classroom discourse (Parrish, 2010); kinesthetic activities; use of concrete materials and math manipulatives; and multiple representations.

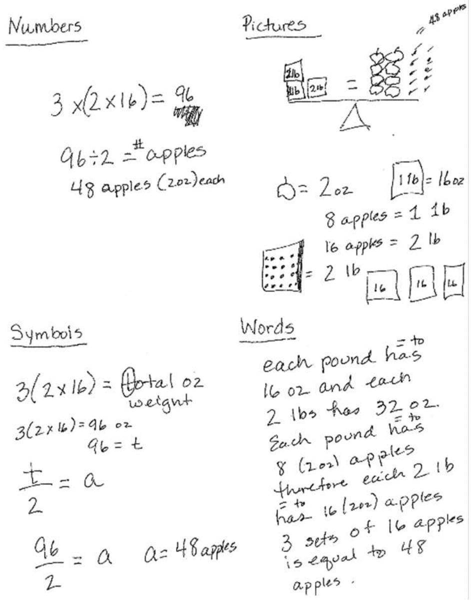

Numbers
3 x(2x16)= 96
96÷2 = # apples
48 apples (2oz)each
Pictures
48 apples
2lb
1lb
2lb
= 2oz
1lb = 16oz
8 apples = 1 lb
16 apples = 2 lb
= 2 lb
16
16
16
Symbols
3(2x16) = total oz weight
3(2x16)=96 oz
96 = t
t/2 = a
96/2 = a
a= 48apples
Words
each pound has = to 16 oz and each 2 lbs has 32 oz.
Each pound has = to 8 (2oz) apples
therefore each 2 lb has = to 16 (2oz) apples
3 sets of 16 apples is equal to 48 apples.

Figure 3 - Multiple

representations of a solution

Throughout each institute, teachers created and maintained their own notebooks (Fig. 3) as a record of their mathematical learning, as a reference to call upon in Lesson Study, and as a model for the student notebooks implemented in classrooms. Thus, teachers mastered the logistical elements of notebook practice, and PD providers modeled the use of graphic organizers, advance organizers, and diagrams enhancing academic language development.

Each Summer, teachers ,and principals participated in 48 hours of intensive PD provided by CSUSB faculty and RCOE staff. Pre-K educators received 24 hours of Summer PD, joining selected plenary sessions and breakouts focused on developmentally appropriate content.

Activity 3. Lesson Study (Goals 2 & 3). This strategy builds upon the LEAs' existing practice of working in Professional Learning Communities to design Units of Study implementing the CCSS-M. TEEM teachers worked in cross-site, grade-level teams to develop CCSS-M-aligned research lessons and assess their impact on ELs by engaging in 36 hours of lesson study per year, facilitated by CSUSB faculty and RCOE staff. Each team produced two research lessons in each project year and research lessons were videotaped for teacher learning. Preschool educators were introduced to lesson study by joining with Kindergarten lesson study teams in Year 1. In later years they formed their own lesson study team.

Activity 4. Professional learning for school leaders (Goal 4). Principal's support is key to the success of any school improvement project. Principals joined teachers in TEEM Summer Institutes, and in two-hour monthly meetings with TEEM staff. In Years 1 and 2, they participated in two facilitated half-day sessions of instructional rounds. Principals worked alongside teachers in lesson study in later years when teams were established in their practice.

Key research questions:

(RQ1) After TEEM was implemented in classrooms, what was the effect on the mean math achievement of English Learner students (ELs) compared to the mean math achievement of ELs in classrooms not implementing TEEM?

(RQ2) After TEEM was implemented in classrooms, what was the effect on the mean English language arts achievement of ELs compared to the mean English language arts achievement of ELs in classrooms not implementing TEEM?

(RQ3) After TEEM was implemented in classrooms, what was the effect on the K-2nd grade ELs' understanding of math concepts to the understanding of math concepts of K-2nd grade ELs in classrooms not implementing TEEM?

(RQ4) Did TEEM have an effect on teachers' knowledge and beliefs about teaching English Learners compared to teachers in the business-as-usual condition?

(RQ5) Did TEEM have an effect on teachers' math instruction provided to English Learners compared to teachers in the business-as-usual condition?

Study Design: The study utilizes a within-school quasi-experimental design.

Study Sample: The study sample consisted of Treatment and Comparison Students and Teachers at five elementary and two middle schools located in two small rural districts in Southern California. The primary sample of interest was English Learner students who encompass between 20% and 27% of all students who attend the target school districts, though the general population was utilized in the analysis in certain circumstances when the sample size of EL students was too small to produce valid results.

Statistical Analysis: Data analyses were conducted using Hierarchical Linear Modeling (HLM) for four of the outcome measures. Analyses consisted of an intervention versus comparison, 2-level model with students nested in classroom. Treating teachers as a nesting unit is justified and appropriate because teachers were the program participation unit and direct exposure to the program is presented at the teacher level. A random effects model was chosen to allow for greater generalizability and interpretation of school level factors. Below, as an example, we specify the statistical model that will estimate the impact of TEEM on students' math achievement as measured by their performance on the Smarter Balanced mathematics assessment.

Model specifications

This is a two-level hierarchical linear model where the treatment indicator appears in level-2, the classroom level.

Level-1 Model: Student Level

$$Y_{ij} = \beta_{0j} + \beta_{1j}(Y^*_{ij}) + \sum_{m=3}^{M} \beta_{mj} x_{mij} + \varepsilon_{ij}$$

where

Y_{ij} is posttest score (e.g., Smarter Balanced math score) for student in classroom j;

β_{0j} is the conditional mean posttest score for control students in classroom j,

Y^*_{ij} is the pretest score for student i in classroom j;

β_{1j} is the average pretest slope for students in classroom j;

x_{mij} are M additional covariates representing demographic characteristics of student in classroom j (e.g. special ed designation, gender, free/reduced lunch), dummies to represent grade-level of students (e.g. 4th, 5th, 6th grades), or other student-level covariates.

β_{mj} are M coefficients corresponding to student-level demographic covariates; and

ε_{ij} is the random effect representing the difference between student ij's score and the predicted mean score for in classroom j. These residual effects are assumed normally distributed with mean 0 and variance σ^2.

Level-2 Model: Classroom Level

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(T_j) + \sum_{q}^{Q} \gamma_{0q} W_{qj} + \mu_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{mj} = \gamma_{m0}$$

where

T_j = 1 if classroom j is an intervention in classroom j, and 0 if control;

γ_{00} is the conditional classroom-level mean for control in classrooms;

γ_{01} is the treatment effect, i.e. the difference between treatment and control classroom conditional means;

γ_{oq} are Q coefficients corresponding to classroom-level covariates ;

μ_{0j} is the deviation of classroom j's mean from the grand mean, conditional on covariates - this effect is assumed normally distributed with mean 0 and variance τ^2

Measures:

- *Student*:
  - Smarter Balanced Mathematics Assessment (grade 3-6)
  - Smarter Balanced English Language Arts Assessment (grade 3-6)
  - Assessing Math Concepts (grades K-2)

- *Teacher*:
  - Mathematics instruction using a Classroom Observation Checklist
  - Teacher Survey
  - Mathematics knowledge – Learning Mathematics for Teaching
  - Math Teaching Efficacy and Expectancy Beliefs Instrument

Student Measures

Assessing Math Concepts:

Assessing Math Concepts is a Kathy Richardson (2011) assessment intended to evaluate early childhood knowledge of math concepts. There are multiple sections, including Counting Objects, Changing Numbers, Number Arrangements, Ten Frames, 2-Digit Addition and Subtraction, and Grouping Tens.

Smarter Balanced Assessment Consortium (SBAC) - ELA and Math:

The SBAC assessment is a computer adaptive standardized measure based on the Common Core Standards. It was given for the first time (other than field testing) in Spring of 2015.

Teacher Measures

Math Teaching Efficacy and Expectancy Beliefs Instrument:

The survey is a 33-item self-report measure intended to evaluate a teacher's self-efficacy and expectancy beliefs related to their ability to teach math. Items include the following, "When a student commits an error in math, I am able to diagnose his/her conceptual errors" and, "I am comfortable helping my English Learners gain conceptual understanding of mathematics." Each statement was rated on a 5-pt. scale from Strongly Disagree to Strongly Agree. The reliability and validity of this instrument is well established. Specifics can be found here: http://ncnaep.rcoe.appstate.edu/sites/ncnaep.rcoe.appstate.edu/files/EnochsSmithBeliefsSurvey2000.pdf. Cronbach's alpha ranged from .77 to .88 (Enochs, Smith, & Huinker, 2000).

Learning Mathematics for Teaching (LMT):

The LMT is a teacher math assessment that measures not only content knowledge but teaching ability in the subject of math. For example, a test item might show a potential error a student has made and ask the teacher to identify the error. The LMT is a robust assessment with good reliability and validity. The primary study conducted by the assessment creators at the University of Michigan on the reliability and validity of the instrument is documented and can be found at http://www.umich.edu/~lmtweb/files/hillshillingball.pdf. IRT reliabilities range from .71 to .84 scales (Hill, Schilling, & Ball, 2004).

Formative Assessment Survey:

The formative assessment survey is a brief 8-item online survey intended to measure the ways in which teachers are using formative assessment (e.g., to facilitate prior learning, adjust teaching strategies, facilitate grouping) and with what frequency. The scale is a 5-pt. scale from 1=never to 5=daily. This survey was designed by the project and its evaluator.

Classroom Observation Checklist:

The classroom observation checklist is a measure of teacher classroom practices and student engagement. The observer must evaluate each item on a 4-pt. scale related to level of implementation from 0=not observed at this time to 3=fully implementing.

The table below represents the schedule of data collection efforts over all four years of the program.

Table 3: Data Collection Schedule

	Baseline	Implementation Year 1		Implementation Year 2		Implementation Year 3		Implementation Year 4	
	Spring 2015	Fall 2015	Spring 2016	Fall 2016	Spring 2017	Fall 2017	Spring 2018	Fall 2018	Spring 2019
Smarter Balanced-Math	√		√		√		√		√
Smarter Balanced-ELA	√		√		√		√		√
Learning Mathematics for Teaching (LMT)	√		√		√		√		√
Assessing Math Concepts				√	√	√	√	√	√
Math Teaching Efficacy and Expectancy Beliefs Instrument	√		√		√		√		√
Classroom Observation Checklist	√		√		√		√		√

Results

Fidelity of Implementation

A fidelity of implementation table (See Appendix 2) was created and applied each year of the project. Categories related to program implementation were defined and operationalized, and each year the evaluation team assigned a score based upon fixed criteria to help gauge whether fundamental attributes of the project were applied with fidelity. It was intended to serve two needs. First, as a formative assessment tool to determine whether the project was on track or whether it needed course correction and second, to inform our interpretation of the findings. As a formative assessment tool, the matrix was not as useful as expected, primarily because the elements which scored lower on the scale could not be corrected due to staffing issues. For example, the original intent was to have principals participate in trainings and take assessments as the teachers had. Unfortunately the project was unable to gather consistent compliance from administration.

Components scored on the implementation matrix included teacher participation in trainings, teacher collaboration, classroom observations, principal participation in trainings, and student notebooking. Each component was scored on a 3-point scale from 0 to 2. For example, student notebooking was based on the average of student scores on their notebooks based upon a rubric. If less than 60% of rubric elements were included in the notebook the student would receive a score of 0. If at least 60% to 79% of the elements were included they would receive a score of 1. If 80% of the rubric elements (or better) were included they would receive a score of 2. The average was taken across students.

A perfect score on the fidelity of implementation table across all categories would be a score of 10. The average for all four years was 5.25 out of 10. The project was unable to achieve a high level of implementation throughout the project largely due to staffing issues and lack of compliance. For the most part, the implementation was high at the teacher level, but less so at the principal and student level. The highest levels of implementation appeared to be in years 2 and 3 which is reflected in the outcomes.

Impact Study Findings

The following section provides a summary of the findings for the confirmatory research questions of the impact study which centered on year 3 of the project. Data sources include two student measures and two teacher level measures. The student level measures are the SBA – Math and ELA scores and the teacher level measures are the LMT and the MTEEBI. In year 3, it was hypothesized that students in TEEM classrooms would outperform their peers in comparison classrooms with no TEEM exposure to a statistically significant degree on both the SBA – Math and ELA exams. HLM was used to analyze the data with students nested in classrooms. The difference between TEEM student performance and those in a comparison classroom were not statistically significantly different. The effect size was .17 for Math and .22 for ELA, falling slightly below the What Works Clearinghouse (WWC) threshold of a “substantively important” Hedge’s g of .25.

HLM was also applied to analyze teacher outcomes; nesting teachers within schools. In year 3 (the confirmatory year), TEEM teachers outperformed a comparison group of teachers who did not participate in the project to a statistically significant degree (p = .04) on the LMT – PFA with an effect size of .33 which is above the WWC threshold of .25.

On the MTEEBI, a measure of teacher expectations of their students and self-efficacy, there was also a statistically significant difference between TEEM teachers and comparison teachers (p = .004) and a substantially significant effect size of .36.

Table 4: Hedge's g Effect Sizes for primary outcomes

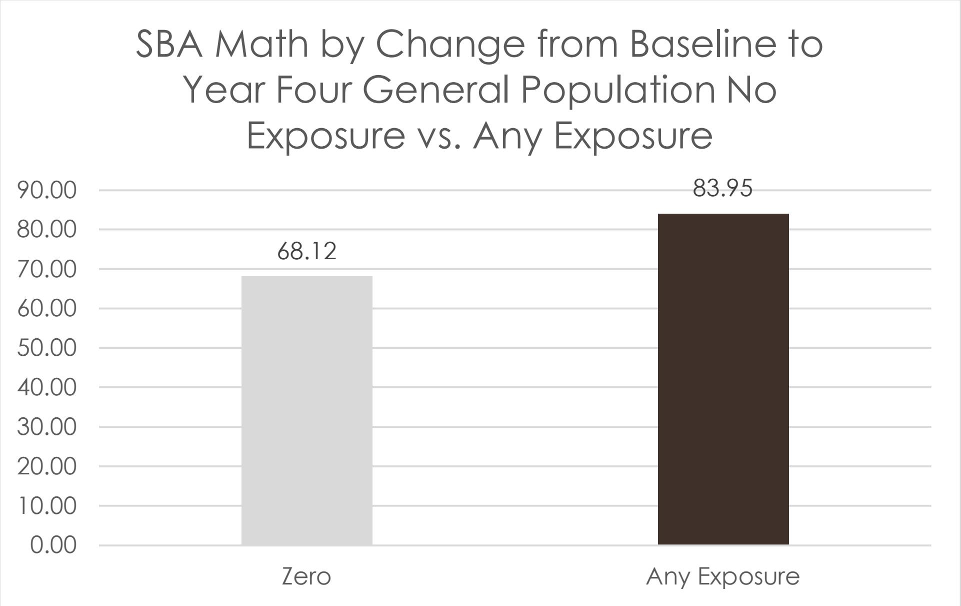
	Effect Size Y1	Effect Size Y2	Effect Size Y3	Effect Size Y4
SBA-Math	.15	.23	.17	.16
SBA-ELA	.09	.38	.22	.20
LMT	.33	.36	.35	.28
MTEEBI	.17	.46	.36	.29

Although the purpose of the TEEM Project was to focus on English Learner students, the project was also interested in the impact of the work on the general student population. In addition, in some cases when disaggregating, the sample size of student subgroups was so small that it precluded analysis. For these reasons, some of the results presented will include all students and some results will be focused on English Learner students specifically.

Smarter-Balanced Assessments – Math

Students took the baseline state level administered SBA – Math at the end of the 2014-2015 school year. The Year 1 post- administration was at the end of the 2015-2016 school year. The Year 2 post- administration was administered in the Spring of the 2016-2017 school year, Year 3 post-administration occurred in the Spring of 2017-2018, and year 4 post-administration occurred Spring of 2018-2019.

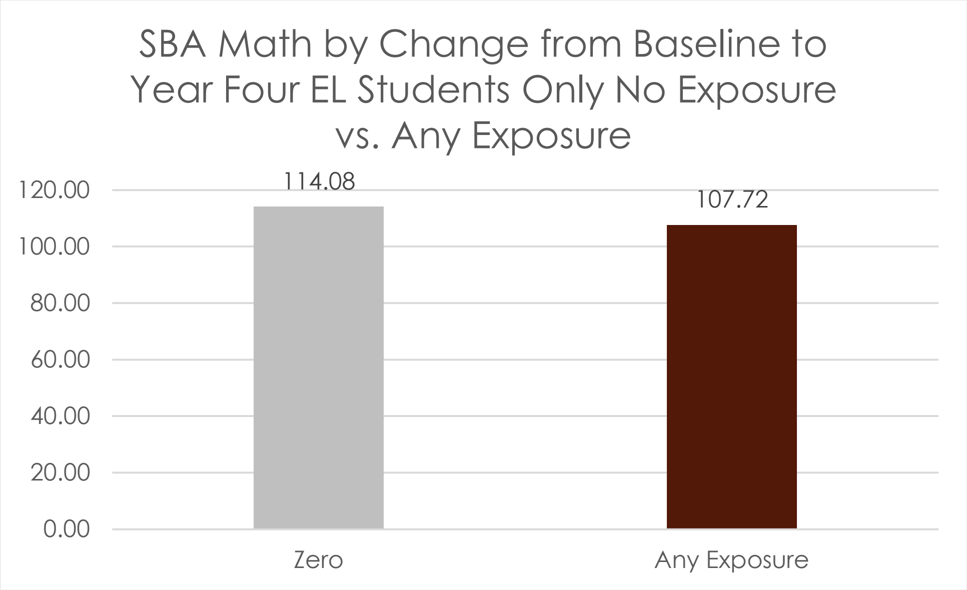
Figure 4: SBA-Math by Years of Exposure General Student Population Average Scale Score Change



* zero exposure n = 2664; any exposure n = 801

The average SBA Math scale score increased for general population students with any exposure to the program at a greater rate relative to students with no exposure to the program. However, statistical testing revealed that this was not quite a statistically significant difference.

Figure 5: SBA – SBA-Math by Years of Exposure English Learner Student Population Average Scale Score change



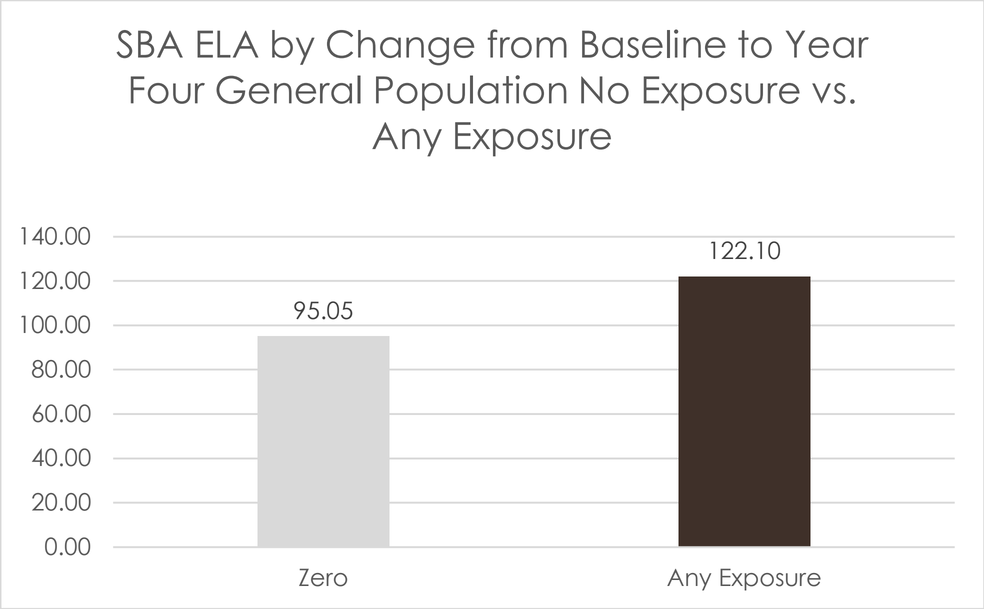
* zero exposure n = 429; any exposure n = 92

In this English Learner specific analysis, Treatment Students demonstrated similar gains relative to the Comparison Group students (zero exposure), with an average increase of 108 points over the course of the study. The difference was not statistically significant.

Smarter-Balanced Assessments – ELA

Figure 6: SBA-Math by Years of Exposure General Student Population Average Scale Score change

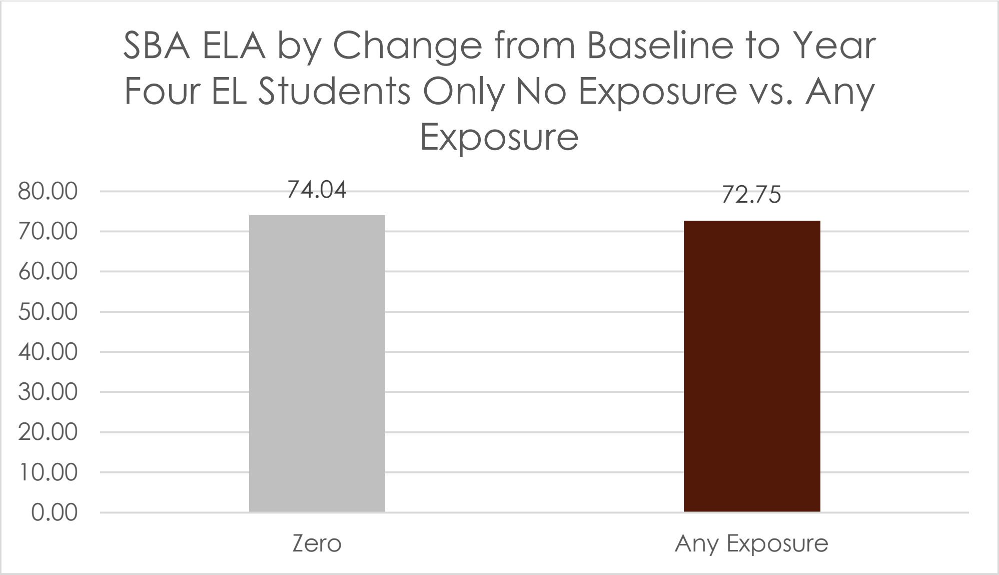
* zero exposure n = 2660; any exposure n = 800



The average SBA ELA scale score of Treatment students increased at a greater rate over the four years of the program, relative to Comparison students. The difference was statistically significant (*F* (1,692) = 6.024, *p* =.014).

Figure 7: SBA – ELA by Years of Exposure English Learner Student Population Average Scale Score Change

* zero exposure n = 63; any exposure n = 129



In this English Learner specific analysis, Treatment Students performed similarly to the Comparison Group students (zero exposure), with an average increase of 73 points over the course of the study. The difference was not statistically significant.

Goal 1: Improve the academic achievement of English Learners.

Benchmark: Participating students will outperform a comparison group in Mathematics and English Language Arts by 5% in each project year, as measured by the Smarter Balanced Assessments.

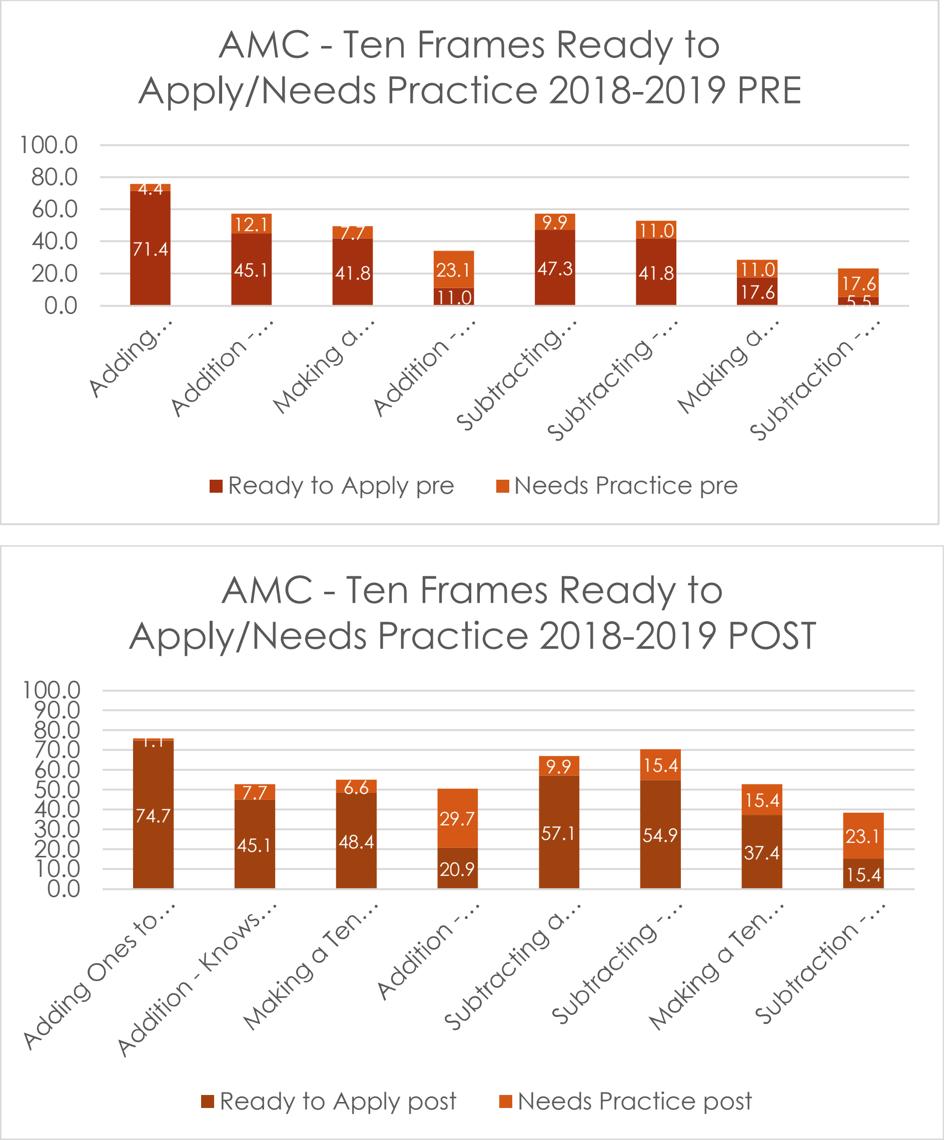
Result: Treatment EL Students, performed similarly to the comparison group, thus, the goal was not met.

Assessing Math Concepts (AMC)

The AMC was the assessment that was used to measure student achievement in grades Pre-Kindergarten through 2nd grade. The AMC was administered twice in Implementation Year 2 - Year 4 at the beginning and end of each year. Only within year comparisons can be conducted with this data because only Treatment students are given the test and most students do not have a Treatment teacher in successive years. The figures to follow reflect the final full year of the project, Y4.

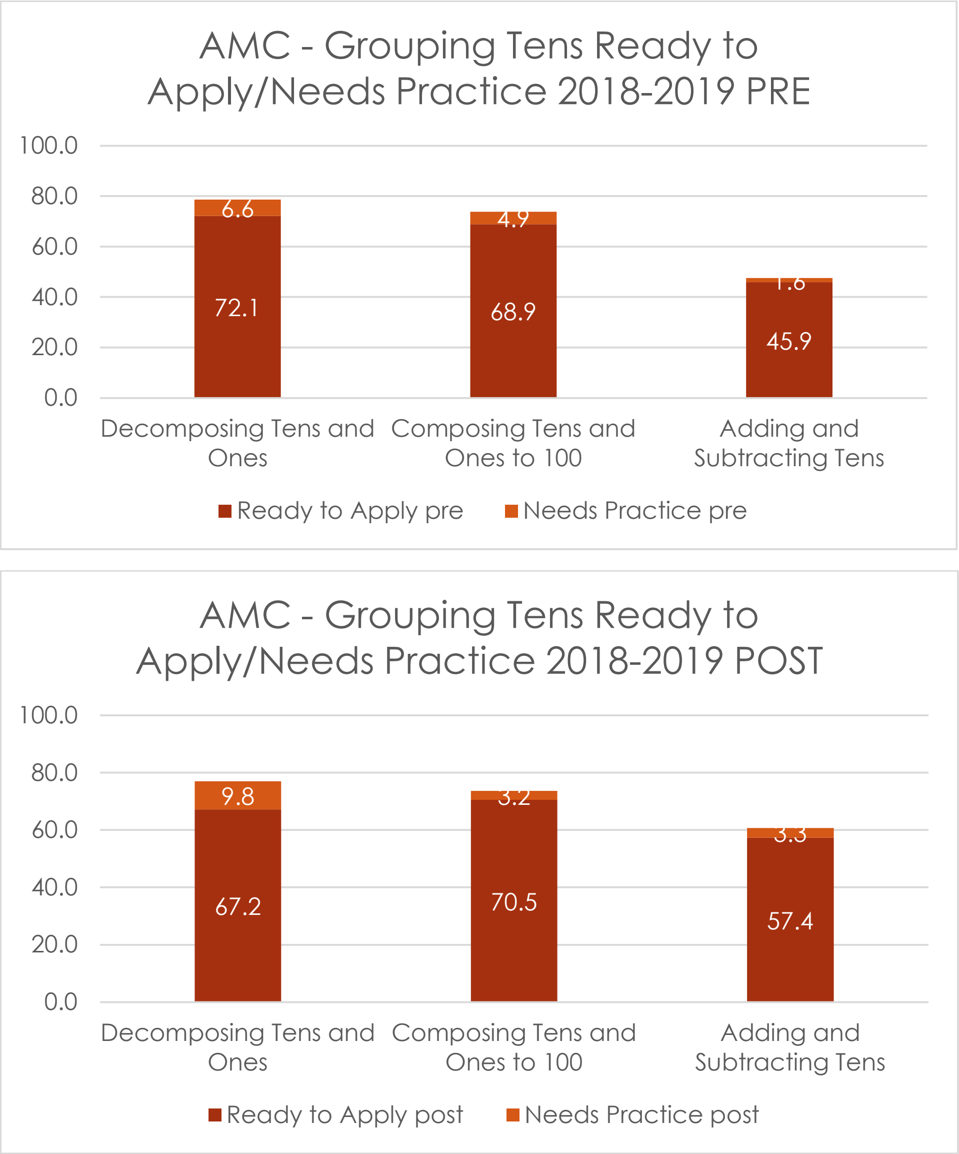
The five sections administered included: Ten Frames (n=91), Grouping Tens (n=61), Number Arrangements (n=169), Two Digit Numbers (n=54), and Counting Objects (n=265). The assessments were administered to students individually, one-on-one, by treatment teachers only. Pre-and post- scores were a matched case file for each section. Results are only provided for students overall, not disaggregated by EL status, due to small sample size.

Figure 8: AMC – Ten Frames Apply/Needs Practice Responses



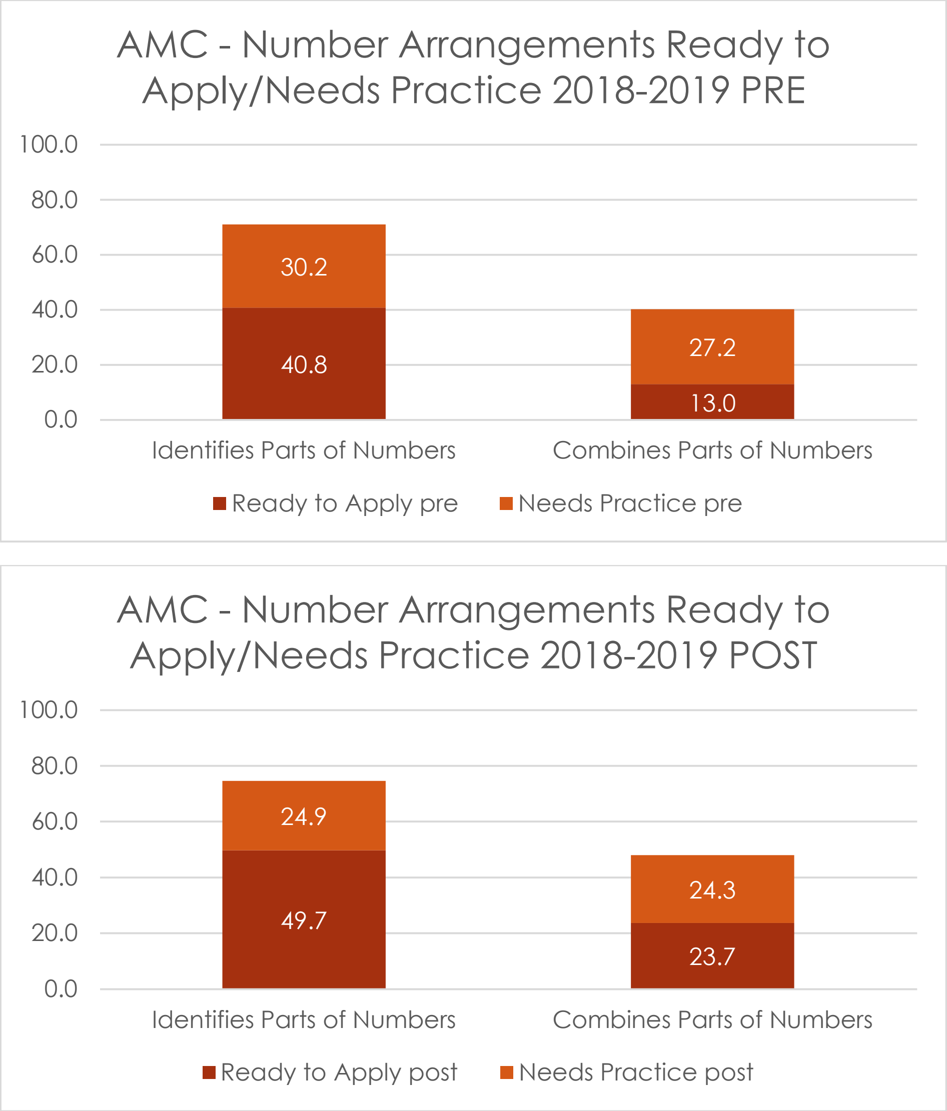
The Ten Frames (n=91) section of the AMC was divided into eight discrete parts including: Adding Ones to a Ten, Knows Parts of Numbers – Addition, Making a Ten and Adding Ones, Recognizing Ten More – Addition, Subtracting a Ten from Ones, Knows Parts of Numbers – Subtracting, Making a Ten and Subtracting Ones, and Recognizing Ten or More – Subtraction. In each case, the percentage of students who earned scores in the top performance level (Ready to Apply) increased from pre-to post with the exception of Addition – Knows Parts of Numbers which remained the same. Making a Ten and Subtracting Ones showed the greatest improvement from pre-to post, increasing from 18% to 37%.

Figure 9: AMC – Grouping Tens Apply/Needs Practice Responses



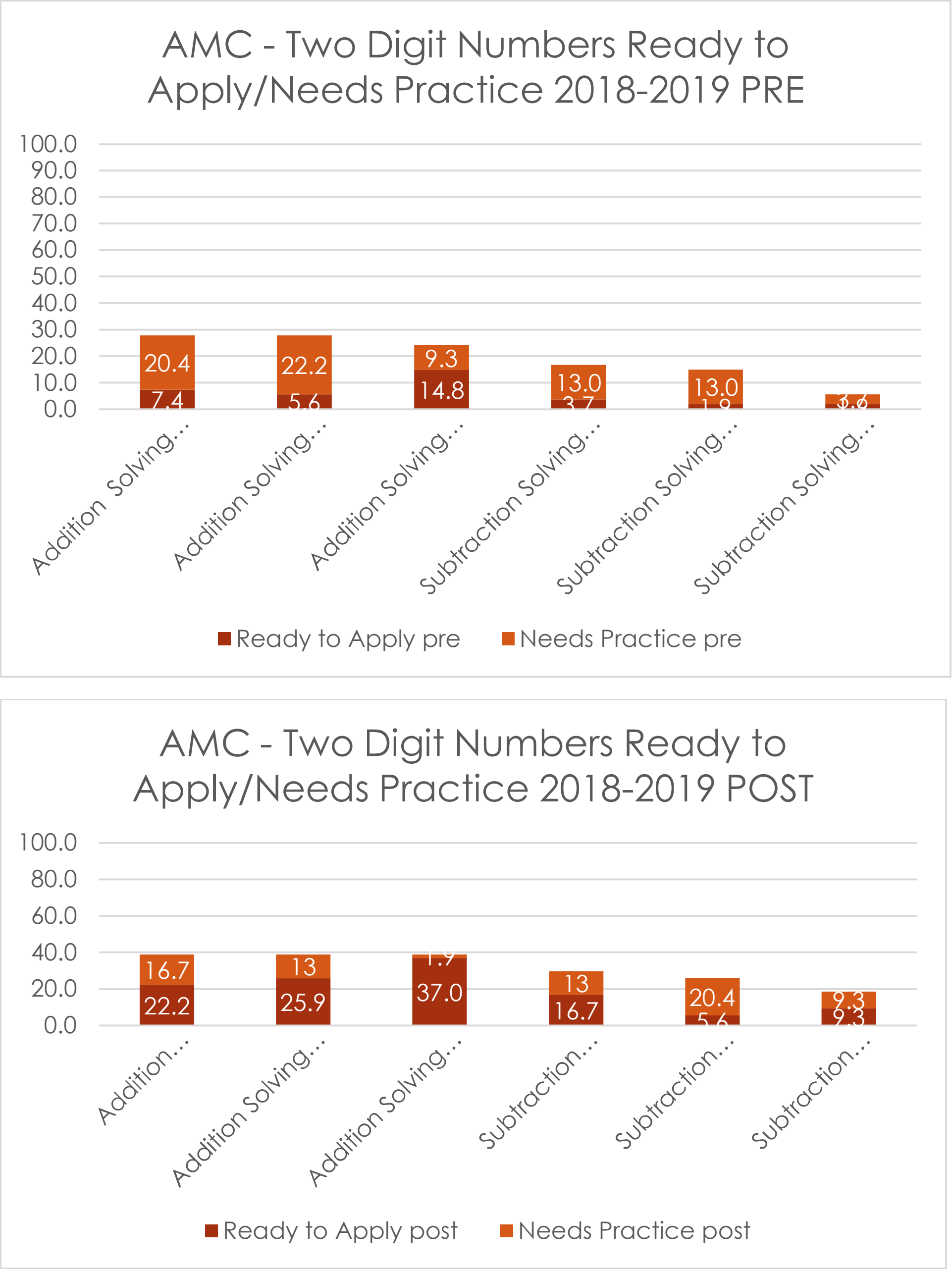
The Grouping Tens (n=61) section of the AMC was divided into three discrete parts including: Decomposing Tens and Ones, Composing Tens and Ones to 100, and Adding and Subtracting Tens. In each case, the percentage of students who earned the top performance level (Ready to Apply) remained relatively the same or increased slightly. Adding and Subtracting Tens increased the most from 46% to 57% over the course of the school year.

Figure 10: AMC – Number Arrangements Apply/Needs Practice Responses



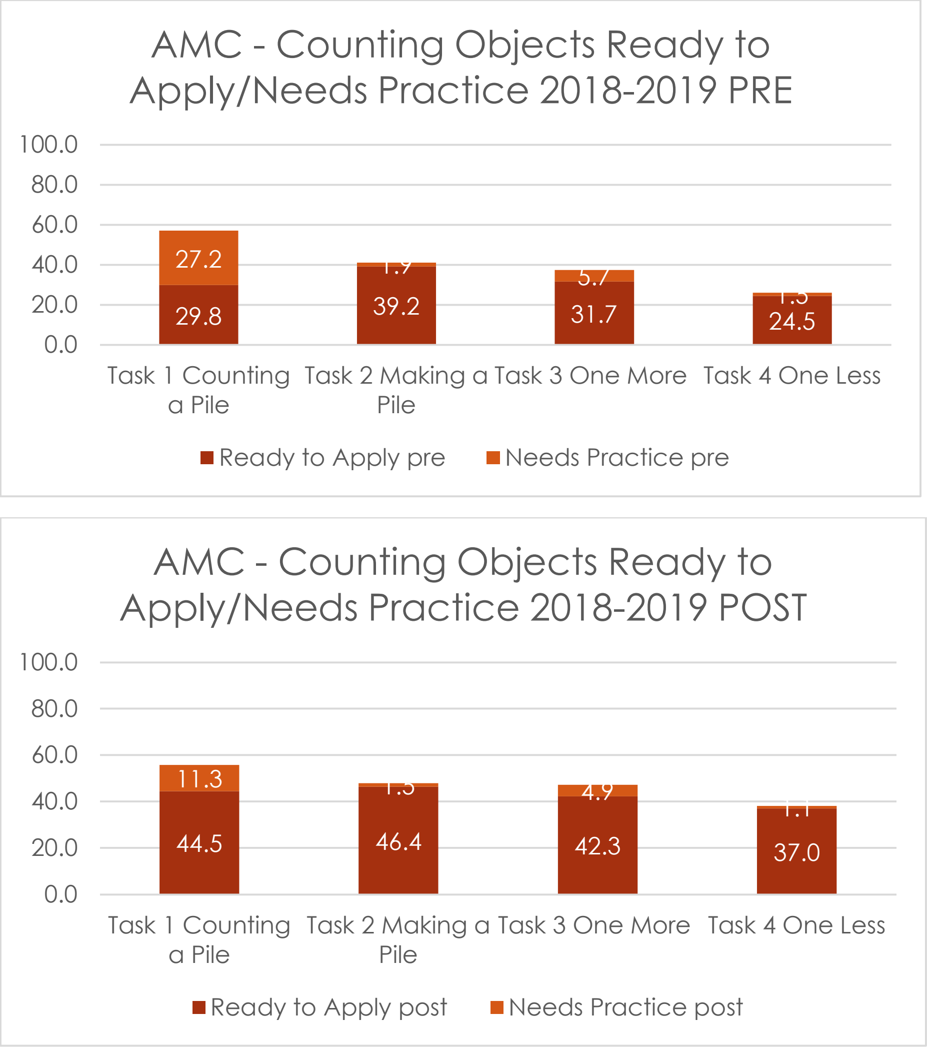
The Number Arrangements (n=169) section of the AMC was divided into two discrete parts including: Identifies Parts of Numbers and Combines Parts of Numbers. In each case, the percentage of students who earned the top performance level (Ready to Apply) increased from pre-to-post. Identifies Parts of Numbers increased by 9 percentage points and Combines Parts of Numbers by 11 percentage points.

Figure 11: AMC – Two Digit Numbers Apply/Needs Practice Responses



The Two Digit Numbers (n=54) section of the AMC was divided into six discrete parts including: Solving Problems with a Model – Addition, Solving Problems without a Model – Addition, Solving Symbolic Problems – Addition, Solving Problems with a Model – Subtraction, Solving Problems without a Model – Subtraction, and Solving Symbolic Problems – Subtraction. At post- students scored highest on Addition Solving Symbolic Problems and Addition Solving Problems with a Model.

Figure 12: AMC – Counting Objects Apply/Needs Practice Responses



The Counting Objects (n=265) section of the AMC was divided into four discrete parts including: Counting a Pile, Making a Pile, One More, and One Less. In each case, the percentage of students who earned the top performance level (Ready to Apply) increased substantially from pre-to post. Students performed best on Making a Pile by post and Counting a Pile by post.

Goal 1: Improve the academic achievement of students.

Benchmark: Participating students will improve their performance by 10% on the Assessing Math Concepts Interviews each year. This goal required amending due to the inability to collect a Comparison sample. The time, cost, and compliance required to complete the interviews were prohibitive. Thus, the goal was amended to a within subject's measurement.

Results: Treatment early elementary students who completed the AMC improved their performance (Needs Practice/Ready to Apply) by more than 10% on every section of the interview with pre-and post- measures with the exception of Grouping 10's. Results were as follows: 10 Frames (pre-score = 47.3%; post- score = 57.9%; prct change = 23%); Grouping 10's (pre-score = 66.7%; post- score = 70.5%; prct change = 6%); Counting Objects (pre-score = 40.4%; post- score = 47.3%; prct change = 17%); Number Arrangements (pre-score = 55.6%; post- score = 61.3%; prct change = 10%); 2-Digit Numbers (pre-score = 19.5%; post-score = 31.8%; prct change = 63%).
Thus, this goal was exceeded for all sections apart from Grouping 10's.

Learning Mathematics for Teaching (LMT)
At baseline, teachers were administered three sections of the LMT; Numbers Concepts and Operations (NCOP), Patterns Functions and Algebra (PFA), and Geometry (GEO). At the end of Implementation Year 1, teachers were administered an alternate form of the NCOP. At the end of Year 2 they were administered an alternate form of the GEO and at the end of Year 3 they completed an alternate form of the PFA section of the LMT. Although Algebra was not the focus of training in Year 3, it was the closest assessment given the options. The PFA was administered again in Year 4.

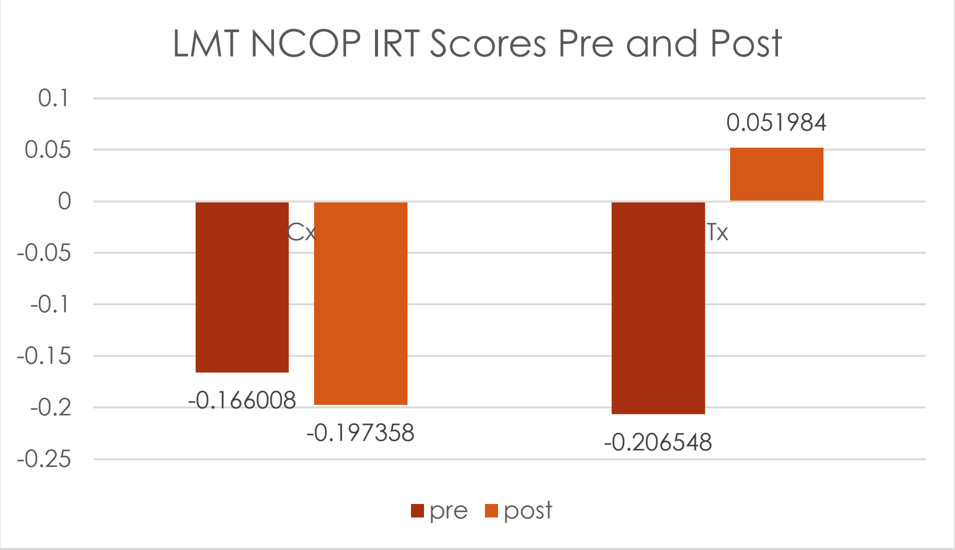
The table and figures to follow show the sample size, minimum, maximum, mean, and standard deviations for the pre- and post-administrations of the NCOP portion of the Learning Mathematics for Teachers assessment. The IRT scores reflect how teachers performed on average at each administration of the assessment. Cx refers to the Comparison group and Tx refers to the treatment group.

Table 5: LMT NCOP Descriptive Statistics by Treatment and Comparison Group

Group		N	Minimum	Maximum	Mean	Std. Deviation
cx	Estimate pre	40	-2.18150	1.23270	-0.16601	0.71903
	Estimate post	40	-1.78800	1.42150	-0.19736	0.71742
	change	40	-1.39000	1.94000	-0.03140	0.80815
	Valid N (listwise)	40				
tx	Estimate pre	62	-2.22000	1.58940	-0.20655	0.74647

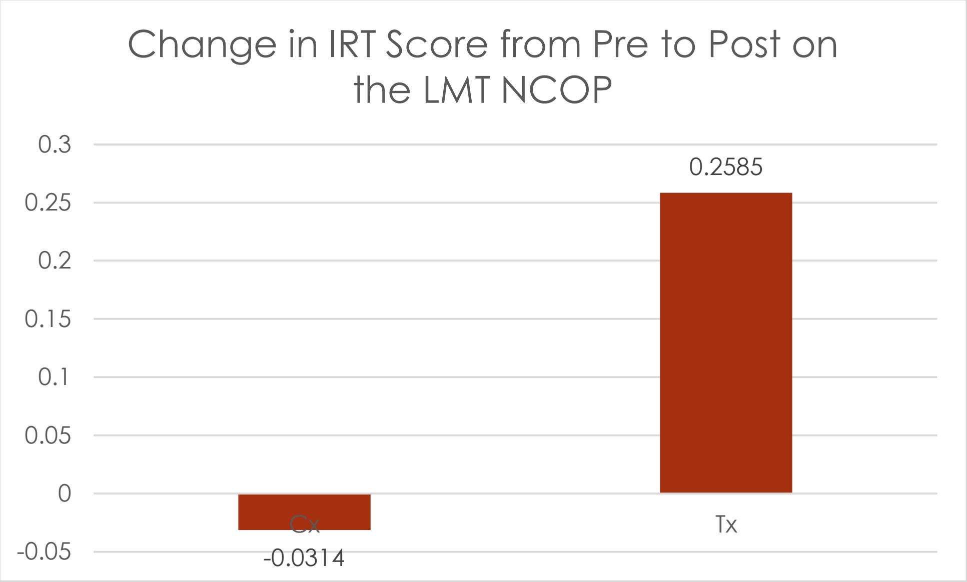
Estimate post	62	-1.95860	1.80910	0.05198	0.80255
change	62	-0.96000	1.62000	0.25850	0.62426
Valid N (listwise)	62				

Figure 13: LMT NCOP IRT Scores by Group



As demonstrated in the graph, Treatment Teachers improved their performance from pre-to post- assessment on the first section of the LMT. This was a marginally statistically significant difference at the *p*=.05 level (*df* = 122; *t* = 1.848; *p* = .06). In contrast, teachers in the Comparison condition decreased from pre- to post-assessment. When comparing the Treatment Teachers and the Comparison Teachers, it should be noted that the Treatment Teachers were performing at a lower level at baseline and improved at a greater rate over the course of the year relative to the Comparison Group.

Figure 14: Change in IRT Score on LMT NCOP by Group



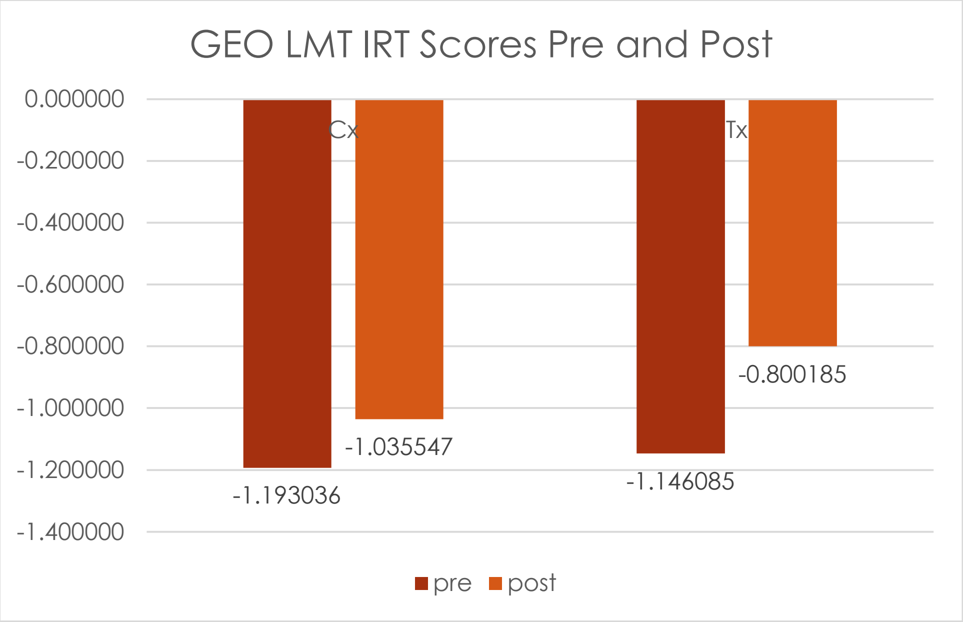
The graph shown above demonstrates the average change in IRT score from pre- to post-assessment on the LMT NCOP. The Treatment Group of teachers performed better on the post-assessment, relative to the pre-assessment, showing growth from pre- to post-. A significance test of the change from pre- to post-measurement between Treatment and Comparison Teachers found that this difference was statistically significant at the *p* = .05 level (*df* = 100; *t* = 2.037; *p* = .04). Treatment Teachers improved their performance on the assessment at a higher rate relative to the Comparison Teachers to a statistically significant degree.

The table and figures to follow show the sample size, minimum, maximum, mean, and standard deviations for the pre- and post-administrations of the GEO portion of the Learning Mathematics for Teachers assessment. The IRT scores reflect how teachers performed on average at each administration of the assessment.

Table 6: LMT GEO Descriptive Statistics by Treatment and Comparison Group

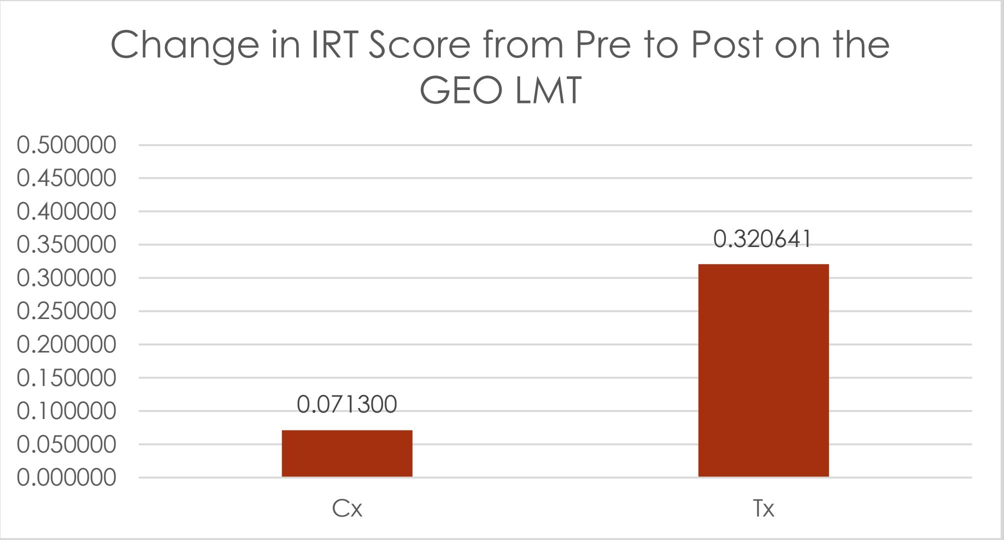
Group		N	Minimum	Maximum	Mean	Std. Deviation
cx	Estimate pre	36	-2.8400	.3949	-1.193036	.6954222
	Estimate post	19	-1.7370	-.1464	-1.035547	.5095198
	change	19	-1.3381	1.2518	.071300	.6247509
	Valid N (listwise)	19				
tx	Estimate pre	40	-2.5048	.9139	-1.146085	.6153893
	Estimate post	34	-2.3100	1.5675	-.800185	.7856406
	change	34	-1.2428	1.8660	.320641	.6682134
	Valid N (listwise)	34				

Figure 15: LMT GEO IRT Scores by Group



As demonstrated in the graph, Treatment Teachers improved their performance from pre-to post- assessment on this section of the LMT. This was a statistically significant difference at the *p*=.05 level (*df* = 72; *t* = 2.123; *p* = .04). Teachers in the Comparison condition also slightly improved from pre- to post-assessment, however, this was not a statistically significant improvement.

Figure 16: Change in IRT Score on LMT GEO by Group



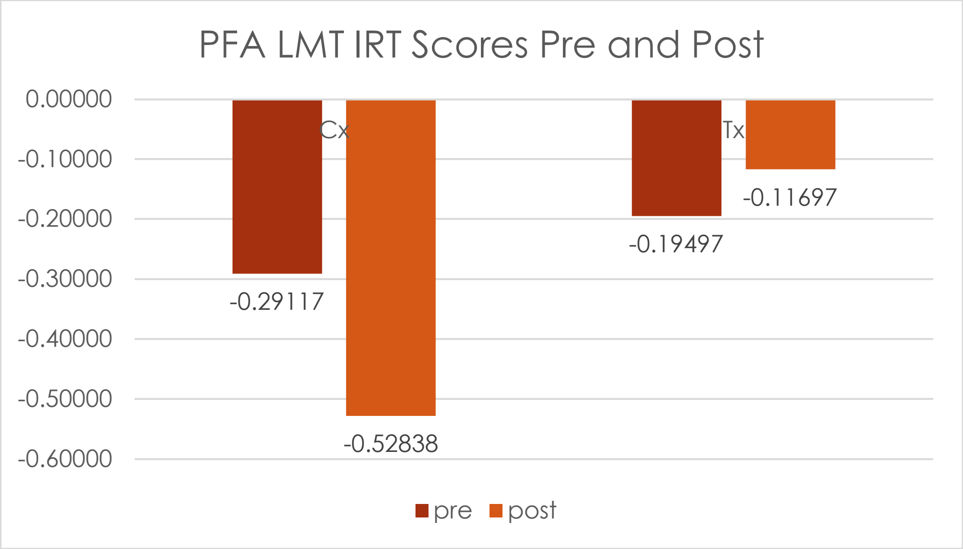
The graph shown in Figure 16 demonstrates the average change in IRT score from pre- to post-assessment on the LMT GEO. The Treatment Group of teachers performed better on the post-assessment, relative to the pre-assessment, showing growth from pre- to post-. A significance test of the change from pre- to post-measurement between Treatment and Comparison Teachers found that this difference was not statistically significant at the *p* = .05 level. The lack of statistical significance may be due to small sample size.

The table and figures to follow show the sample size, minimum, maximum, mean, and standard deviations for the pre- and post-administrations of the PFA portion of the Learning Mathematics for Teachers assessment for Year 3. The IRT scores reflect how teachers performed on average at each administration of the assessment.

Table 7: LMT PFA Year 3 Descriptive Statistics by Treatment and Comparison Group

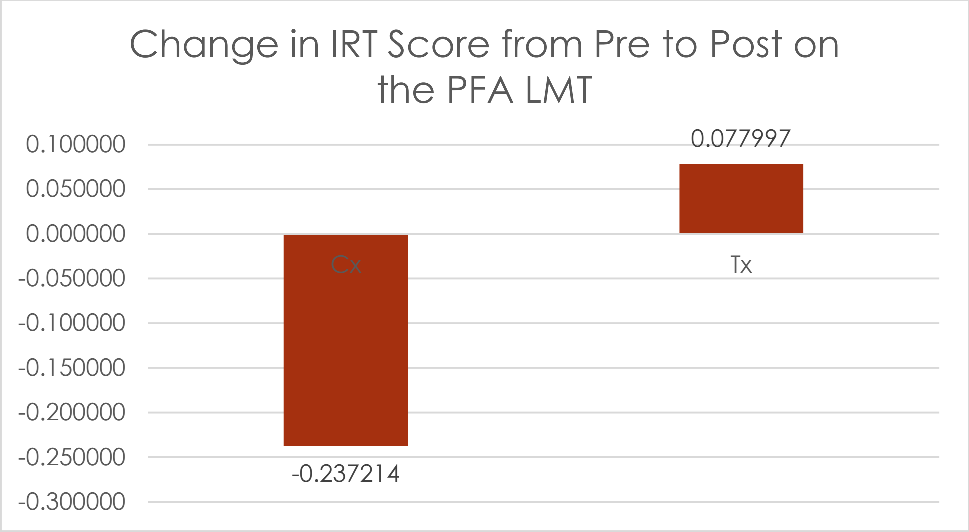
Group		N	Minimum	Maximum	Mean	Std. Deviation
cx	Estimate pre	33	-2.85444	1.52645	-0.29117	0.79037
	Estimate post	33	-2.33033	1.32737	-0.52838	0.91593
	change	33	-1.64532	2.11477	-0.23721	1.04400
	Valid N (listwise)	33				
tx	Estimate pre	42	-1.28158	1.72839	-0.19497	0.75139
	Estimate post	42	-1.51923	1.79305	-0.11697	0.70788
	change	42	-1.91215	1.81421	0.07800	0.84428
	Valid N (listwise)	42				

Figure 17: LMT PFA Year 3 IRT Scores by Group



As demonstrated in the graph, Treatment Teachers improved their performance from pre-to post- assessment and Comparison Teachers declined on this section of the LMT, though this was not statistically significant, likely due to sample size.

Figure 18: Change in IRT Score on LMT PFA Year 3 by Group



The graph demonstrates the average change in IRT score from pre- to post-assessment on the LMT PFA. The Treatment Group of teachers performed better on the post-assessment, relative to the pre-assessment. A significance test of the change from pre- to post-measurement between Treatment and Comparison Teachers found that this difference was not statistically significant at the *p* = .05 level. The lack of statistical significance may be due to small sample size.

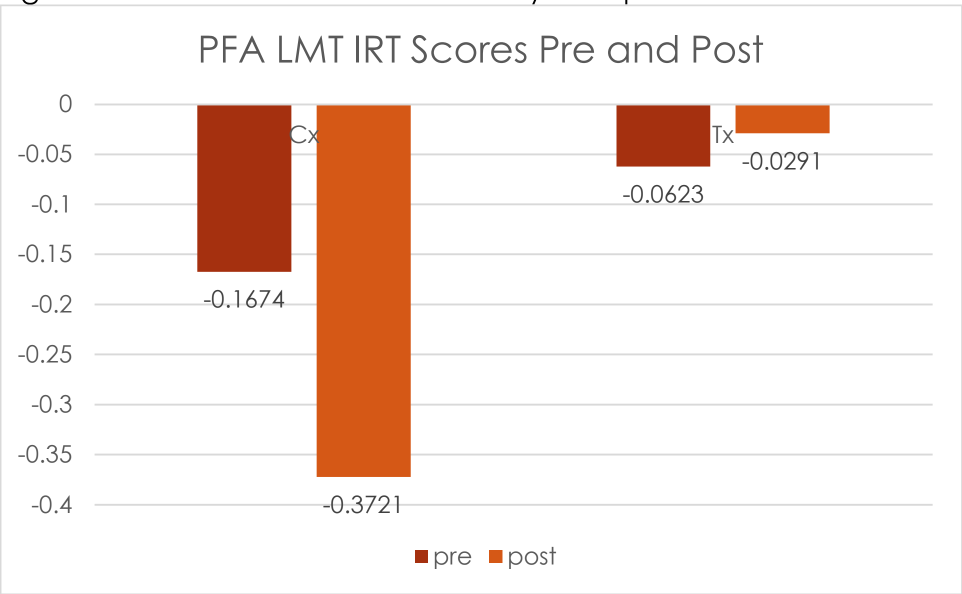
The table and figures to follow show the sample size, minimum, maximum, mean, and standard deviations for the pre- and post-administrations of the PFA portion of the Learning Mathematics for Teachers assessment in Year 4. The IRT scores reflect how teachers performed on average at each administration of the assessment.

Table 8: LMT PFA Year 4 Descriptive Statistics by Treatment and Comparison Group

Descriptive Statistics

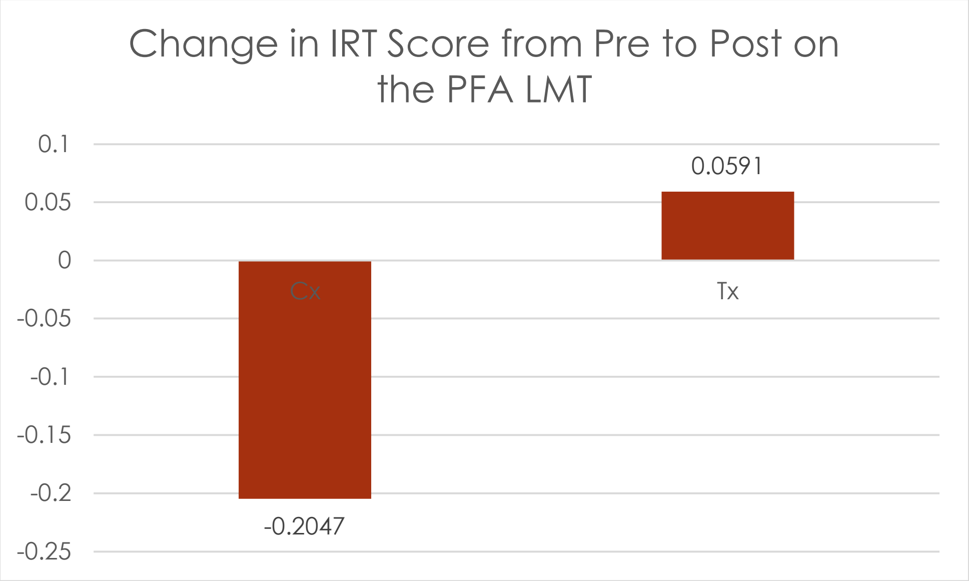
Group		N	Minimum	Maximum	Mean	Std. Deviation
cx	Pre_Score	14	-1.60315000	.57736300	-.1674055929	.68326521575
	Post1_Score	13	-2.1274700	1.3273700	-.430871208	.8865478967
	Post2_Score	14	-1.95297000	.97930800	-.3721421029	.82518009535
	Pre_Post2_ScoreChange	14	-2.1535370000	1.652496000	-.204736510	.95021862
	Valid N (listwise)	13				
tx	Pre_Score	33	-1.21408000	1.72839000	-.0623278764	.71377423947
	Post1_Score	33	-1.3846700	1.7930500	-.075672064	.7531531091
	Post2_Score	35	-1.80529000	2.24067000	-.0291206309	.80487203647
	Pre_Post2_ScoreChange	33	-2.100332000	1.684529000	.059083446	.83854499
	Valid N (listwise)	32				

Figure 19: LMT PFA Year 4 IRT Scores by Group



As demonstrated in the graph, Treatment Teachers improved their performance from pre-to post- assessment slightly on the PFA Section of the LMT. A significance test of the change from pre- to post-measurement between Treatment and Comparison Teachers found that this difference was not statistically significant at the *p* = .05 level likely due to small sample size.

Figure 20: Change in IRT Score on LMT PFA Year 4 by Group



The graph shown demonstrates the average change in IRT score from baseline to Year 4 assessment on the LMT PFA. The Treatment Group of teachers performed better on the post-assessment, relative to the baseline, demonstrating growth. A significance test of the change from pre- to post-measurement between Treatment and Comparison Teachers found that this difference was not statistically significant at the *p* = .05 level. The lack of statistical significance may be due to small sample size.

Goal 2: Improve teacher knowledge and beliefs for teaching mathematics to English Learners.

Benchmark: Participating teachers will outperform comparison teachers in assessments of mathematical content knowledge by 10% in each project year, as measured by the Learning Mathematics for Teaching assessment.

Results: Treatment Teachers outperformed the Comparison Teachers on the NCOP (112% diff), GEO (127% diff), PFA Year 3 (159%), and PFA Year 4 (129%) portions of the LMT, thus exceeding the goal by a good margin.

Math Teaching Efficacy and Expectancy Beliefs Instrument (MTEEBI)

The MTEEBI was administered to teachers, twice in Implementation Year 1; once at the beginning of the school year, and again, following one year of implementation. The survey has been administered at the end of each school year since the first administration. In Year 1, Fifty-one Treatment Teachers completed the MTEEBI. In Year 2, 33 Treatment teachers completed the survey. In Year 3, 62 treatment teachers completed the survey and in Year 4, 34 completed the survey. Survey completion was online using survey data collection software. The results presented in tabular form, represent the percentage of teachers who indicated one of the top two categories (agree and strongly agree) in response to each statement. Results have been divided into two tables, one for self-reported Confidence or Efficacy in their teaching ability, and one for teacher Expectancies for student abilities and performance. The tables are sorted from greatest to least change from baseline to Year 4 post-survey.

Table 9: Percentage of Agree/Strongly Agree Responses to the MTEEBI on Questions Pertaining to Confidence by Group

Confidence	baseline	post yr1	post yr2	post yr3	post yr4	change
I can teach students to determine on their own which situations require an exact answer and which require an estimate	49.20%	67.20%	63.30%	88.30%	87.50%	38.30%
I am able to make sure my students can use materials to represent problems in multiple ways	62.30%	83.60%	77.40%	90.90%	96.90%	34.60%
I can easily integrate students' strategies and ideas into my math lessons even if they are different from my lesson plan	56.70%	80.30%	76.70%	97.00%	90.60%	33.90%
I know how to prepare students to consider the meanings of units used in different contexts	57.40%	83.60%	80.60%	88.30%	90.60%	33.20%
I can help students learn to work on their own to gather appropriate evidence to support their mathematical ideas	68.90%	81.70%	77.50%	97.00%	96.90%	28.00%
I can help students learn to see relationships between quantities	67.20%	73.80%	90.30%	97.00%	93.80%	26.60%
I understand math concepts well enough to be effective in teaching elementary school math	78.70%	83.60%	90.40%	97.00%	100%	21.30%
I am comfortable allowing my students to make their own approximations or simplifications when approaching a real-life problem	73.80%	85.20%	87.10%	97.00%	93.80%	20.00%
I can teach my students to decompose and re-combine numbers and expressions in different ways depending on the context	73.80%	85.00%	83.90%	87.90%	93.80%	20.00%
I am able to help students from impoverished backgrounds excel in math	70.50%	90.20%	87.10%	90.90%	87.50%	17.00%
I am comfortable letting my students struggle with a problem for which there is no immediately obvious method of solution	78.70%	86.90%	90.40%	84.90%	93.80%	15.10%

I am comfortable helping my English Learners gain conceptual understanding of mathematics	76.70%	80.00%	90.30%	87.80%	90.60%	13.90%
I can develop students' ability to produce mathematics (e g a number sentence, expression or equation) to model their own interpretation of a situation	80.00%	95.00%	96.70%	90.90%	93.80%	13.80%
I am comfortable analyzing and synthesizing different student approaches to a mathematics problem to bring closure to a mathematical discussion	72.10%	77.00%	93.50%	79.90%	84.40%	12.30%
I can teach students to make a habit of asking themselves whether their work makes sense	78.70%	85.20%	93.40%	97.00%	90.30%	11.60%
I can incorporate multiple representations into my lessons to improve student learning	80.00%	90.20%	96.70%	90.90%	90.60%	10.60%
I know how to develop students' ability to use the math they know to solve problems in everyday life	82.00%	83.60%	87.10%	90.90%	90.60%	8.60%
I feel comfortable addressing students' questions about mathematical concepts and ideas	88.50%	85.20%	96.80%	90.90%	96.90%	8.40%
When a student commits an error in math, I am able to diagnose his/her conceptual errors	82.00%	82.00%	96.80%	97.00%	87.50%	5.50%
I know how to prepare students to plan their own approaches to solving problems	86.90%	93.30%	96.80%	96.90%	90.70%	3.80%
I feel comfortable teaching students to understand relationships between concepts of algebra and concepts of arithmetic	70.50%	80.30%	80.70%	87.90%	65.60%	-4.90%

The top 5 responses that increased the most from pre-to post- for the Treatment Group on these questions pertain to expectations in their teaching were:

- *I can teach students to determine on their own which situations require an exact answer and which require an estimate*
- *I am able to make sure my students can use materials to represent problems in multiple ways*
- *I can easily integrate students' strategies and ideas into my math lessons even if they are different from my lesson plan*
- *I know how to prepare students to consider the meanings of units used in different contexts*

- *I can help students learn to work on their own to gather appropriate evidence to support their mathematical ideas*

Table 10: Percentage of Agree/Strongly Agree Responses to the MTEEBI on Questions Pertaining to Expectations by Group

Expectations	baseline	post yr1	post yr2	post yr3	post yr 4	change
Students at my grade level think concretely, and teachers can’t be expected to teach them to work with abstractions in mathematics	3.30%	9.80%	10.00%	9.10%	12.50%	9.20%
When students are given the opportunity to make their own generalizations, they end up more confused than if the teacher teaches the mathematics directly	4.90%	8.20%	0.00%	0.00%	12.50%	7.60%
A teacher can be expected to help a student learn math despite his or her impoverished home environment	93.30%	88.50%	93.40%	87.90%	96.90%	3.60%
Even a very skilled teacher cannot expect English Learners to attempt to understand complex mathematics problems	8.20%	8.20%	6.50%	3.00%	9.40%	1.20%
No matter what the teacher does, students can’t seem to determine when an approximate answer is appropriate	11.50%	4.90%	3.20%	3.00%	12.50%	1.00%
Increased effort in math teaching produces little change in some students' math achievement	9.80%	9.80%	10.00%	6.10%	9.40%	-0.40%
Students who have low motivation for learning math can be turned on to learning by their math teachers	96.70%	93.40%	93.60%	97.00%	93.80%	-2.90%
Regardless of the teacher’s instruction, students won’t use available tools to investigate problems on their own	14.80%	6.60%	16.10%	9.10%	9.40%	-5.40%
Seeing many different approaches to solve one problem confuses many students and hinders their learning	22.00%	3.30%	3.20%	3.00%	15.50%	-6.50%
No matter how skilled the teacher, some students won’t understand what quantities mean, even if they can compute them	23.00%	18.30%	3.20%	3.00%	15.60%	-7.40%
Even a teacher with good math teaching abilities may not help some students learn math	33.30%	19.70%	6.50%	15.60%	25.00%	-8.30%

Even with appropriate instruction, most students rarely consider whether their math work makes sense	32.80%	21.30%	6.50%	18.20%	22.60%	-10.20%

The 5 responses that changed the most from pre-to post- for the Treatment Group on questions regarding teacher expectations of student performance were the following. Agreement with each of these statements decreased between 5 to 10 percentage points from baseline to end of Year 4:

- *Even with appropriate instruction, most students rarely consider whether their math work makes sense*
- *Even a teacher with good math teaching abilities may not help some students learn math*
- *No matter how skilled the teacher, some students won't understand what quantities mean, even if they can compute them*
- *Seeing many different approaches to solve one problem confuses many students and hinders their learning*
- *Regardless of the teacher's instruction, students won't use available tools to investigate problems on their own*

There were decreases in several negative beliefs between the baseline and Year 3. There was a bit of a regression in Year 4 regarding expectations of students as can be observed in Table 10.

Goal 2: Improve teacher knowledge and beliefs for teaching mathematics to English Learners.

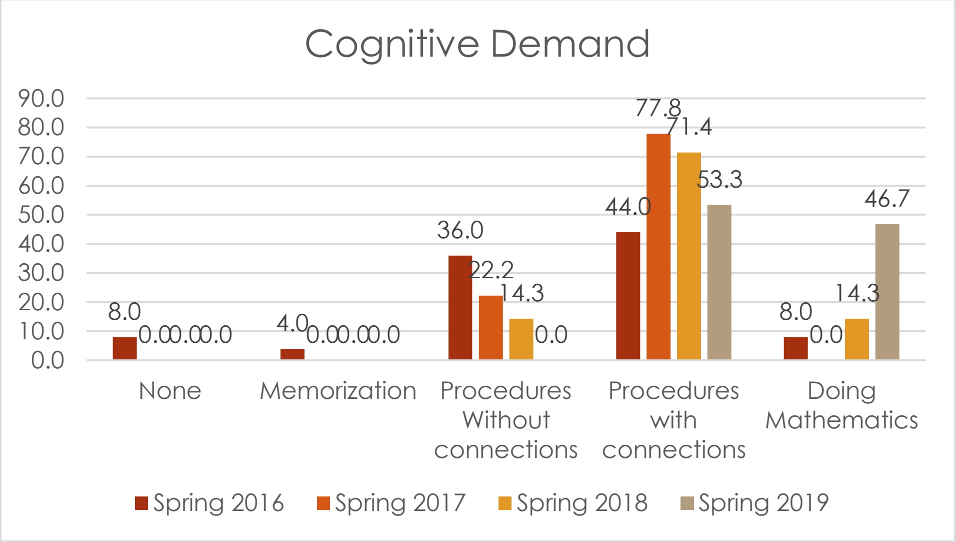
Benchmark: Teachers will show a growth percentage of 5% each year with a greater percentage relative to a comparison group on the Mathematics Efficacy and Outcome Expectancy Beliefs Instrument (MTEEBI).

Results: Treatment Teachers decreased their MTEEBI score from 76% to 72% (prct change = -5.3%) relative to Comparison Teachers who decreased from 78% to 69% (prct change = -12.8%), thus this goal was not met.

Classroom Observations

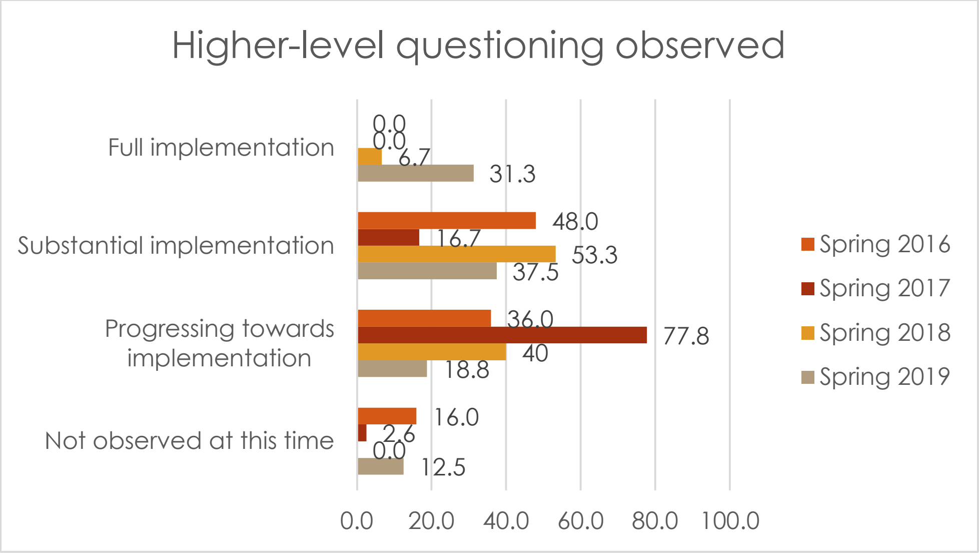
A random sample of 25 TEEM teachers was selected for classroom visits in the first year. Expert observers (who had achieved consensus in the first two years of the project with a Chronbach's Alpha of .80 or better) visited each of the classrooms in Spring of 2016 (n=25), Spring of 2017 (n=18), Spring of 2018 (n=15), and Spring of 2019 (n=16) and completed a 24-item tool with questions pertaining to the following domains: Quality of Mathematics and Cognitive Demand, Language and Discourse, and Classroom Culture.

Figure 21 – Cognitive demand



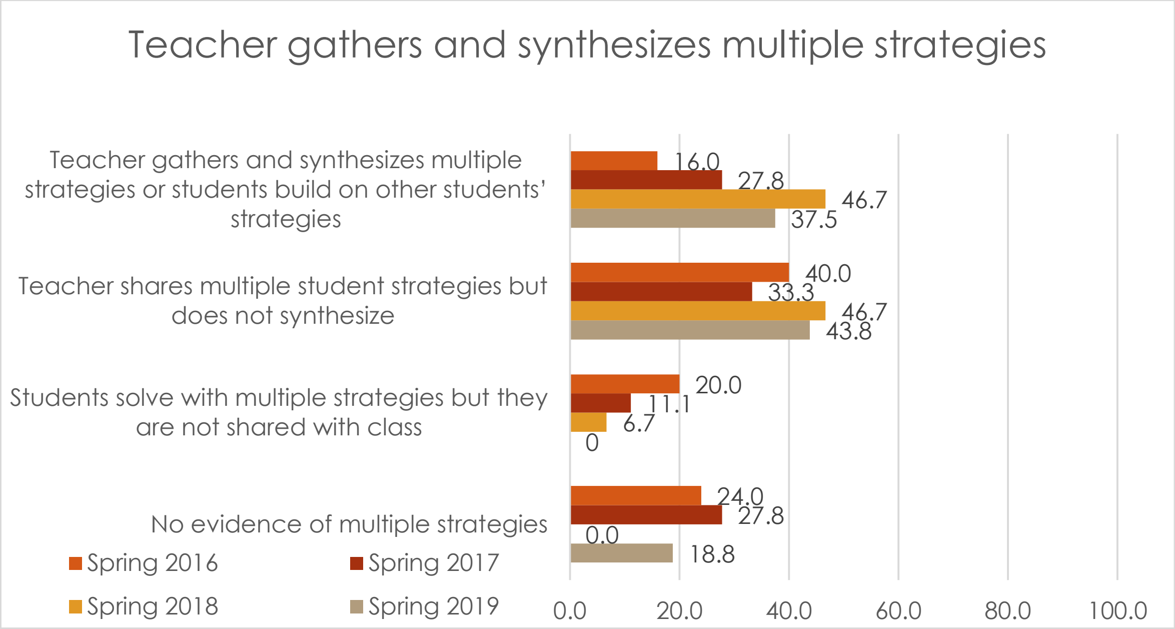
Evidence of doing mathematics increased substantially in year 4. Procedures with connections was at its highest level in year 2 and has declined each year since then.

Figure 22 – Higher-level questioning



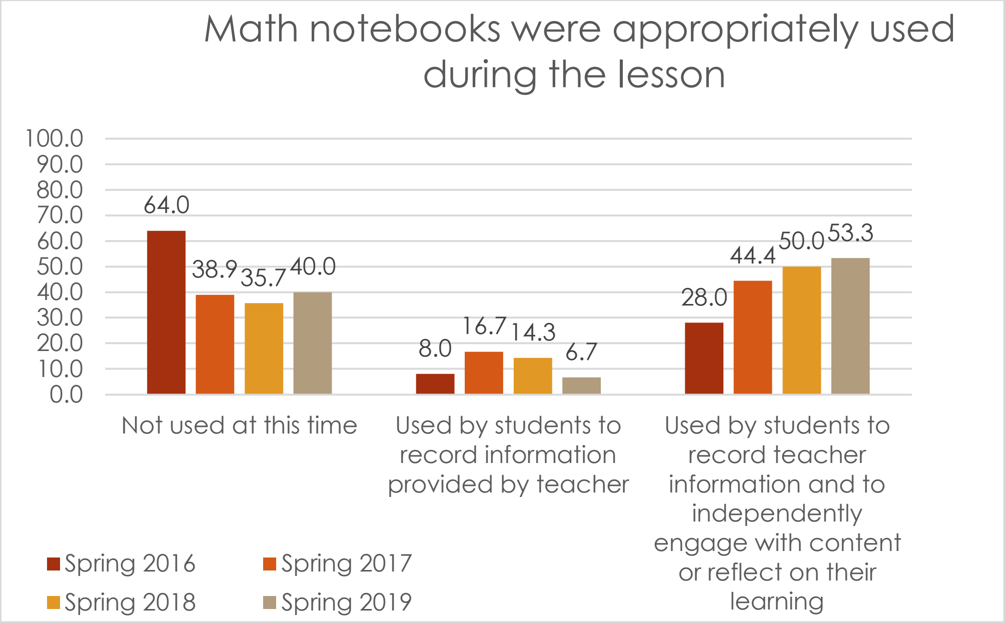
Higher-level questioning was observed slightly less than half of the time at initial measurement. By time 3, 60% of the classrooms observed were either progressing towards or at substantial implementation.

Figure 23 – Teacher uses multiple strategies



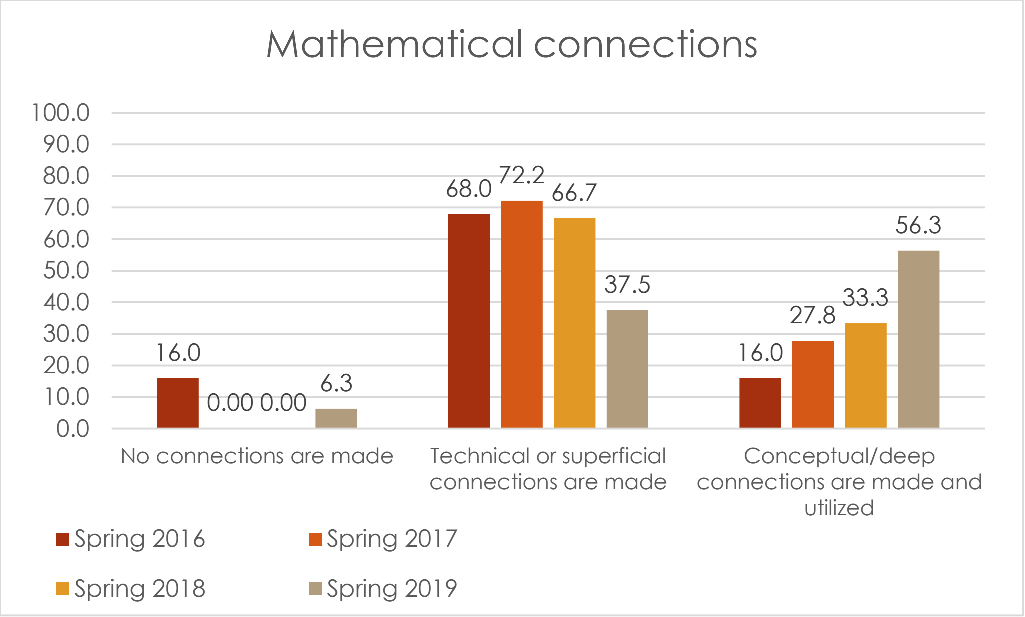
The highest level of teacher synthesis of multiple strategies was only observed 16% of the time in this first observation and increased substantially to 47% of the time by time 3. By time 4, nearly 70% were at substantially or fully implementing. Teacher gathers and synthesizes multiple strategies or students build on other students' strategies began at 16% observation in year 1. It rose steadily to nearly 47% at time 3 and then declined to 38% at time 4.

Figure 24 – Math notebooks used appropriately



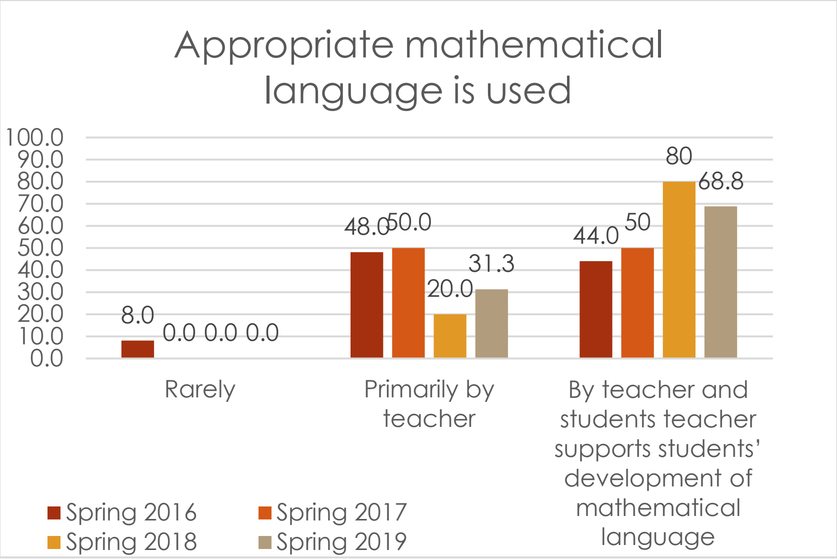
At time 1, use of math notebooks was not widely observed (64% were not using them). By time 2 this percentage had gone down to 39% and 36% at time 3. At time 1, 28% of the time students were observed using the notebooks to record teacher information and engage with content. This went up by 22 percentage points from the first to third administration points. By year 4, usage to record teacher information was at its highest rate of 53%.

Figure 25 – Mathematical connections



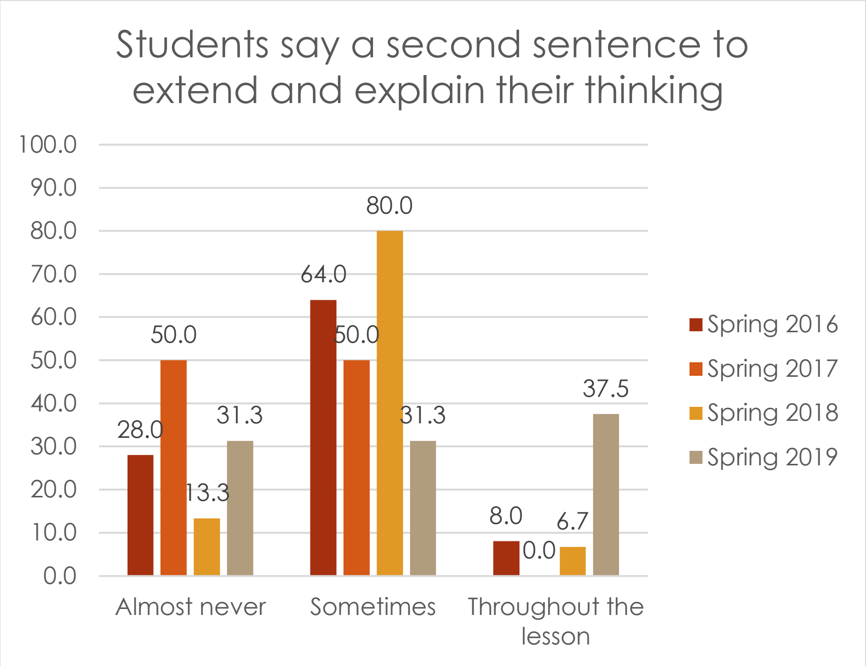
The observation tool required the observers to indicate the level of mathematical connections observed during the observation period. Conceptual or deep connections were only observed about 16% of the time at time 1 and doubled by time 3. By time 4, conceptual and deep connections were at their highest at 56%.

Figure 26: Appropriate mathematical language



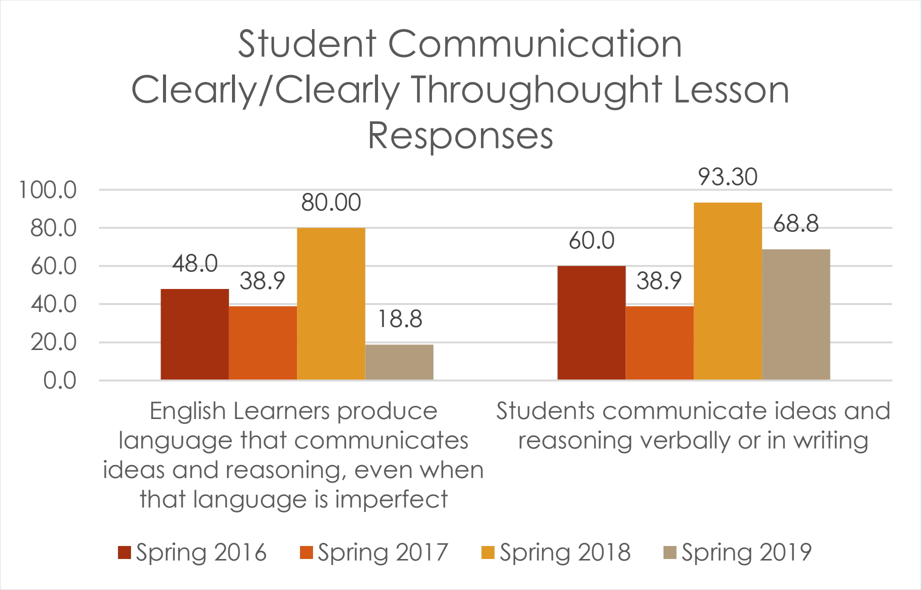
Use of appropriate mathematical language grew considerably over the course of the grant from 44% at time 1 to 80% at time 3, then dipped down to 69% by time 4.

Figure 27 – Students extend and explain thinking



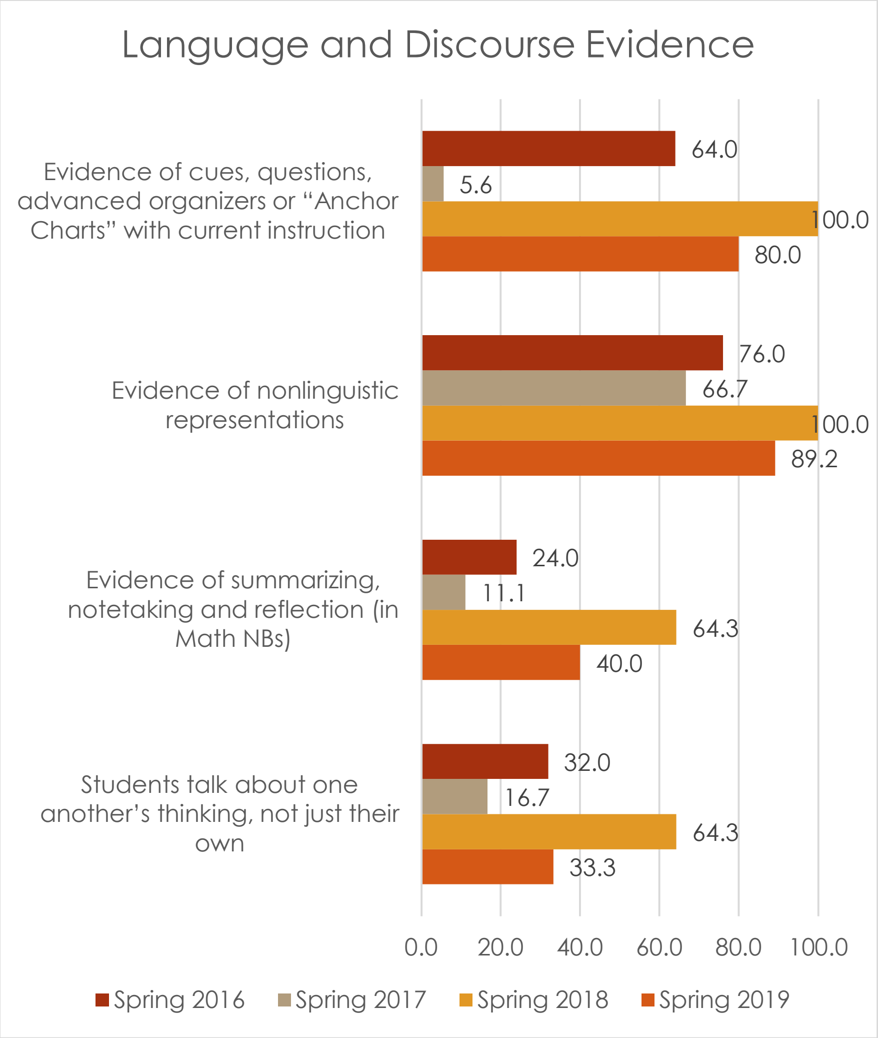
Students were observed uttering a second sentence to extend and explain their thinking "sometimes", which had the greatest frequency of observation at 64% at time 1. These figures increased from time 1 to time 3. By time 3, the percentage had gone up to 80%. By time 4, this was observed 69% of the time either sometimes or throughout the lesson.

Figure 28 – Student communication



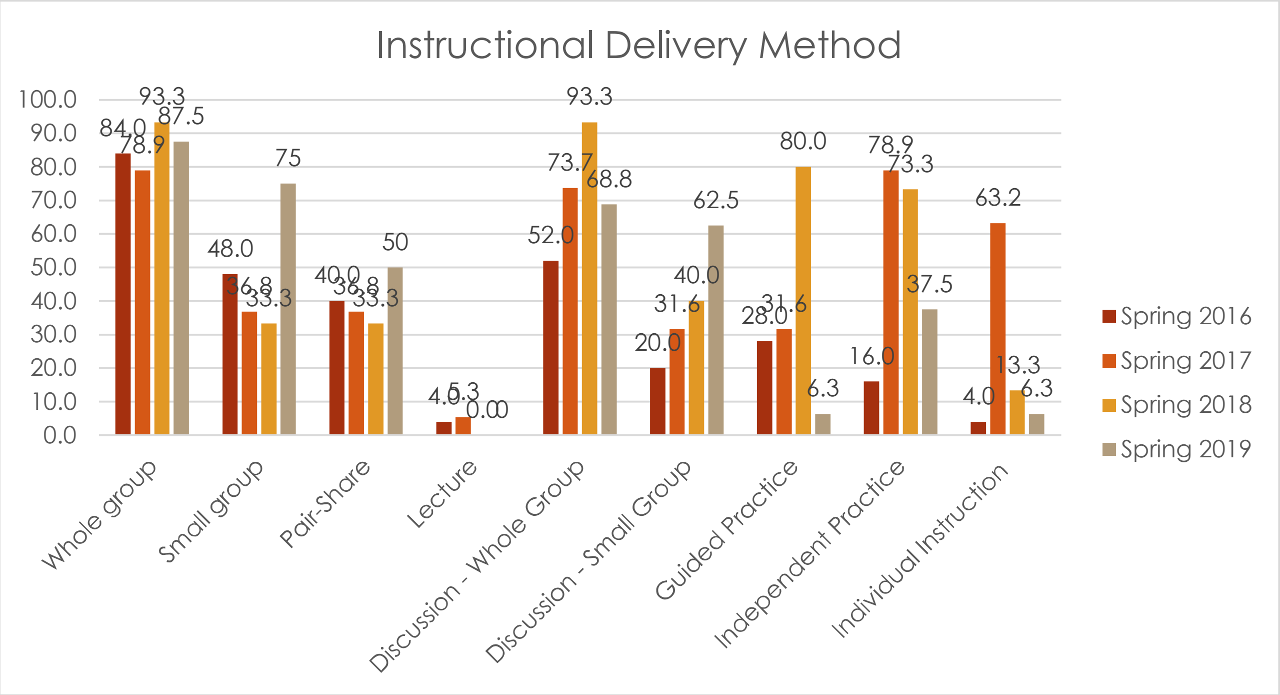
Students were observed communicating their ideas and reasoning verbally or in writing clearly about 60% of the time. This percentage declined in time 2 and was then observed 93% of the time by time 3. English Learners did so about half the time observed. These figures declined slightly by time 2 and then went up to 80% by time 3. By time 4 they had declined considerably to 19%.

Figure 29 – Language and discourse



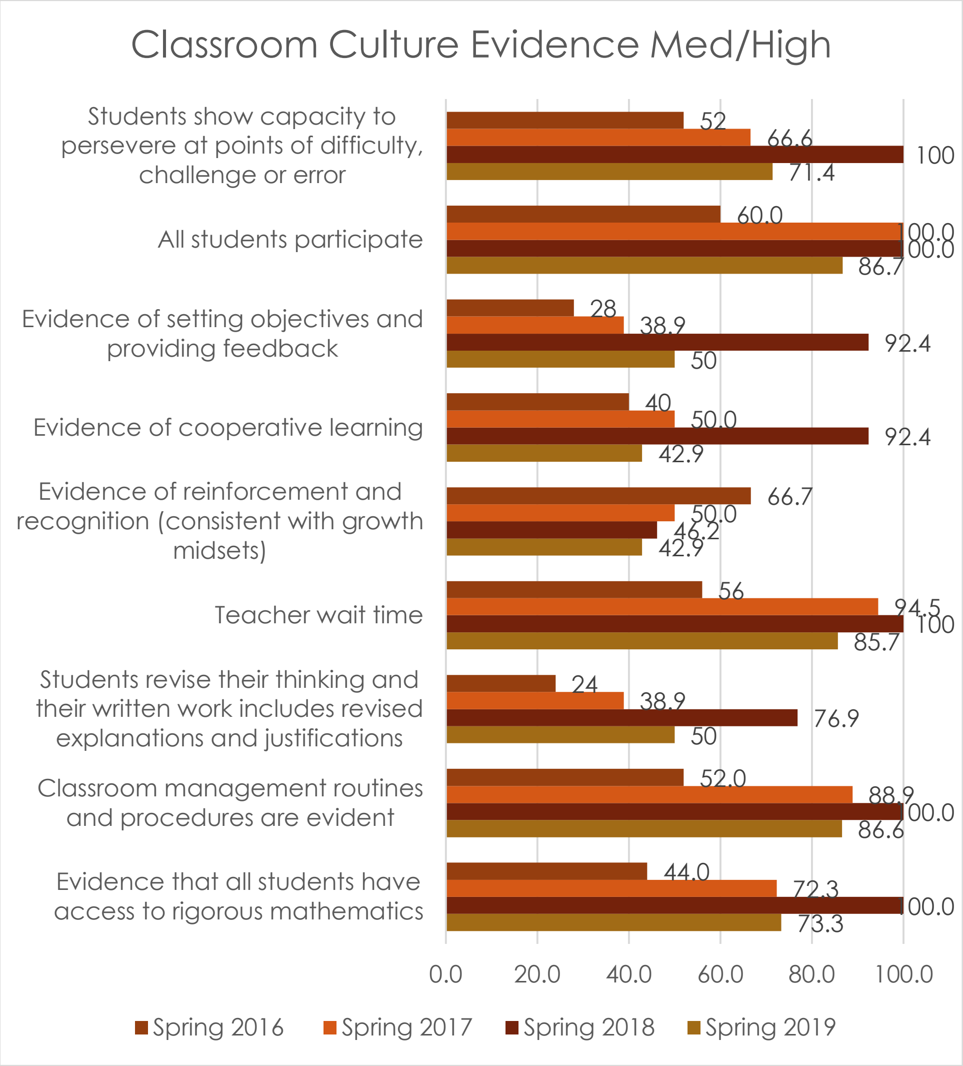
At time 1, Anchor Charts were evident 64% of the time, nonlinguistic representations were visible 76% of the time at time 1. Both of these percentages went up to 100% by time 3 and declined by 10 to 20 percentage points by time 4. Notetaking was evident 24% of the time at time 1. This percentage increased to 64% of the time by time 3 and declined to 40% by time 4. Students were observed talking about one another's thinking approximately 32% of the time at time 1 and 64% of the time at time 3. By time 4 they were at 33%.

Figure 30 – Instructional delivery



At time 1, the most prevalent form of instructional delivery was Whole Group at 84%, followed by Discussion – Whole Group (52%), and Small Group (48%). Individual Instruction and Lecture were the two methods least likely to be observed during the observation period at 4% of the time. By time 3, there were increases in Whole Group Discussion (93%) and Guided Practice (80%). By time 4, Small Group and Whole Group were observed most frequently.

Figure 31 – Classroom culture



The items observed under the domain of Classroom Culture with the highest levels of evidence were: All Students Participate, Classroom Management Routines and Procedures Evident, and Teacher Wait Time. The items with lower levels of evidence include: Students Revise Their Thinking and Work, Setting Objectives and Providing Feedback, and Cooperative Learning. The rankings did not change substantially from time 1 to time 3, but they did increase a great deal from time 1 to time 3. By time 4, most of these areas were not observed with as much frequency as they had been at time 3.

Goal 3: Improve the teaching of math to English Learners and all students.

Benchmark: Teachers will show 25% improvement by project close, or maintain a score of 80%, as measured by a classroom observation checklist.

Result: An overall implementation score was calculated using 25 of the scaled items on the observation checklist. In 2016 the percentage of partial/full implementation observed was 64.9%. In 2017 the percentage remained relatively static at 63.0% of partial/full implementation. In 2018 the percentage went up to 70.7%. In 2019 the percentage declined to 60%. Thus, this goal was not met.

Math Notebooks
A random sample of Treatment Student notebooks was generated by the evaluator. Treatment Teachers were given a list of six students in their classes and asked to bring in at least three notebooks for review. Raters used a notebooking assessment tool which consisted of 18 items. There were 66 notebooks reviewed in total at baseline (2016), 99 at Year 2 post, and 96 at Years 3 and 4 post.

Table 11: Notebooking Assessment Results

	Substantial/Full Implementation				
	baseline Yr 1	post Yr 2	post Yr 3	post Yr 4	change baseline to Yr 4
Evidence of summarizing or reflection (could be quizzes or "exit tickets")	1.5	14.1	63	22.1	20.6
Pages/sections consistently numbered	39.4	46.5	67.5	59.3	19.9
Table of Contents	12.1	18.2	65.1	31.6	19.5
Evidence that Standards for Mathematical Practice are consistently implemented	0	30.3	58.6	18.8	18.8
Evidence of comprehension strategies (ex. "noticing and wondering")	7.6	22.3	74.1	25.1	17.5
Input page shows evidence of concept development: notetaking, recording, graphic organizers and/or practice	68.2	74.7	80.6	85.1	16.9
Literacy strategies infused in lessons (ex: appropriate academic vocabulary development, sentence frames)	0	32.3	61.9	16.6	16.6
Output page shows evidence of higher order thinking, synthesis and/or reflection	27.3	35.4	61.8	37.5	10.2

Evidence of Cooperative Learning activities in the Math Notebook (record of pair or group work in input or output pages)	0	6	37.1	6.2	6.2
Evidence multiple representations (ex: pictures, numbers, symbols, words)	43.9	66.7	64.5	48.9	5
Evidence of consistency in notebook content across class	87.9	99	96.9	89.5	1.6
Evidence of qualitative feedback	0	10.1	45.5	1.1	1.1
Evidence of generating and testing hypotheses/predictions	0	0	44.4	0	0
Defined roles & accountability for cooperative learning teams	0	4	0	0	0
Evidence of points/totals recorded (ex: in Table of Contents)	4.5	10.2	71.4	4.3	-0.2
Signature or stamp used regularly to indicate completion of a task	33.3	32.3	58.6	32.7	-0.6
Math Note Book is used daily as main learning tool	62.1	66.7	74	45.2	-16.9
Designated Input and Output pages consistently used	56.1	52.6	72	39	-17.1

The results of the notebooking review reflect a great deal of variability from year to year. Compliance tended to peak in year 3 for most variables and decline in year 4. As the grant period winds down, teachers may be in the process of determining what elements to maintain and what elements could be discontinued.

The 5 categories that changed the most from pre-to post- on the notebooking rubric were:

- *Evidence of summarizing or reflection (could be quizzes or "exit tickets")*
- *Pages/sections consistently numbered*
- *Table of Contents*
- *Evidence that Standards for Mathematical Practice are consistently implemented*
- *Evidence of comprehension strategies (ex. "noticing and wondering")*

Formative Assessment Survey

The formative assessment survey is a brief 8-item online survey intended to measure the ways in which teachers are using formative assessment (e.g., to facilitate prior learning, adjust teaching strategies, facilitate grouping) and with what frequency. The scale is a 5-pt. scale from 1=never to 5=daily.

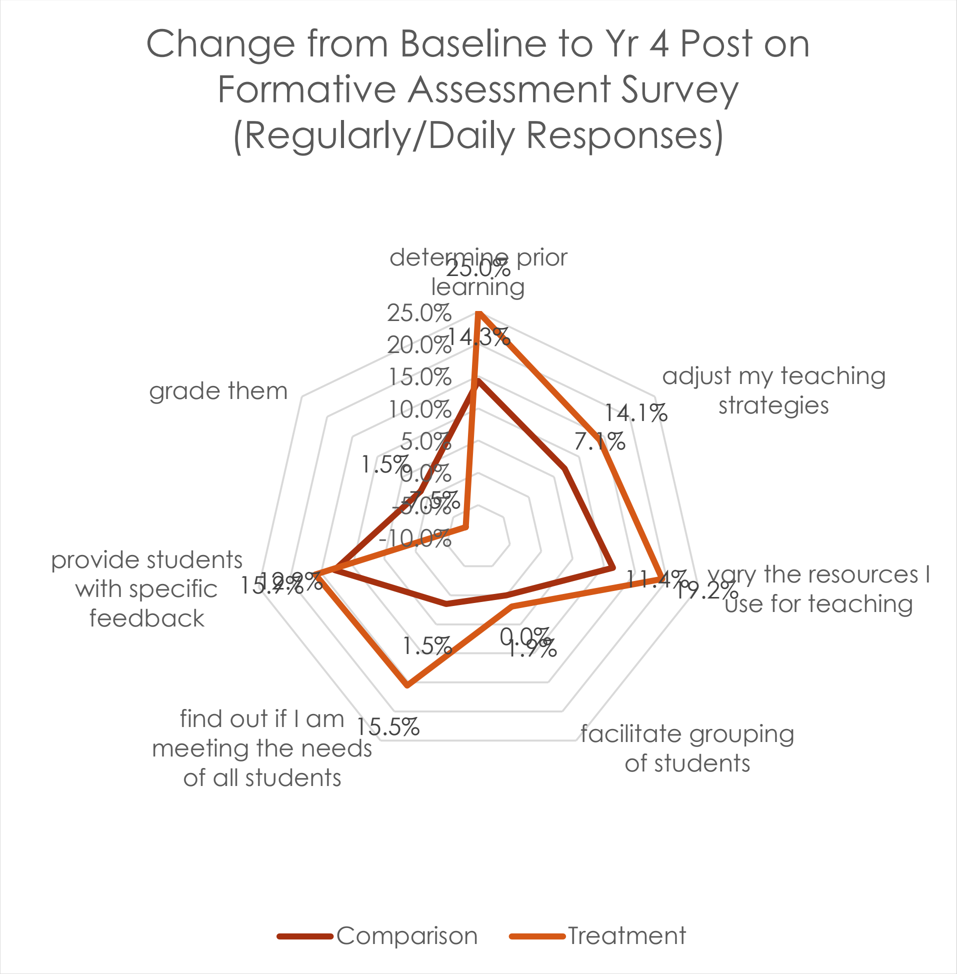
The Formative Assessment Survey was administered to teachers five times. The first administration was at the beginning of the school year in 2015 (baseline), again, following one year of implementation (Year 1 post), a third time at the end of implementation year 2 (Year 2 post), a fourth time at the end of implementation year 3 (Year 3 post), and a fifth time at the end of year 4 (Year 4 post). Sixty-one Treatment Teachers, and 35 Comparison Teachers completed the Formative Assessment Survey at baseline and Year 1 post. At Year 2 post, 33 Treatment Teachers and 16 Comparison Teachers have completed the survey. At Year 3 post, 62 Treatment Teachers and 41 Comparison Teachers have completed the survey. At Year 4 post, 34 Treatment Teachers and 14 Comparison Teachers completed the survey. The survey was completed online using survey data collection software. The results presented below in the table represent the change in the percentage of teachers who indicated one of the top two categories (regularly/daily) in response to how frequently they implement the following formative assessment strategies sorted from greatest change to least (Treatment).

Table 12: Formative Assessment Survey Results Change From time 1 to time 4

	Comparison	Treatment
determine prior learning	14.26%	24.98%
vary the resources I use for teaching	11.41%	19.23%
provide students with specific feedback	12.90%	15.95%
find out if I am meeting the needs of all students	1.47%	15.49%
adjust my teaching strategies	7.09%	14.13%
facilitate grouping of students	-0.01%	1.93%
grade them	1.46%	-7.52%

The Treatment Teachers demonstrated higher growth percentages from baseline to post-on the rate at which they used formative assessment to determine prior learning, provided students with specific feedback, and varied the resources they tended to use for teaching.

Figure 32: Change in Formative Assessment Survey Responses



The spider graph depicts the percentage of change from baseline to Year 4 by teachers who indicated that they regularly or daily utilize the listed formative assessment strategies. The percentage of Treatment teachers who determine prior learning increased by 25 percentage points from baseline to Year 4 post. This was in contrast to the Comparison teachers who increased less than 15 percentage points relative to the Treatment teachers on that variable. 14% of Treatment teachers indicated that they regularly or daily adjust their teaching strategies relative to half that percentage for Comparison teachers.

Goal 3: Improve the teaching of math to English Learners and all students.

Benchmark: Teachers will improve the quality and frequency of classroom formative assessment by 10% each year or 30% by project close, as measured by teacher surveys and notebook assessment rubrics.

Results: A scale score was calculated for the notebooking rubric at the end of each year. At the end of Year 1, the average score was 34.6%, at the end of Year 2 the average score was 34.8%, and at the end of Year 3, the average score was 48.3%. At the end of Year 4, the average score was 35.5%. From baseline to end of Year 4 growth was at 3%, which did not meet expectations for this metric.

A formative assessment score was calculated each year. At baseline, teachers earned a score of 70.4%. At the end of Year 1 they earned a 75.7%. At the end of Year 2 the score earned was 75.4% and at the end of Year 3 they earned a score of 72.0%. By the end of Year 4 they earned a score of 80.1%. Thus, from baseline to time 4, there was a 14% increase, thus exceeding the 10% increase goal.

Unfortunately, we were unable to collect the data we had planned to test Goal 4 pertaining to principal and leadership training. We began with a sample size of 7 and due to scheduling and in some cases, lack of cooperation, we were unable to collect a sufficient sample.

Key Findings

Observations

Observations of Procedures with Connections declined each year since time 2 when it was at its highest. Evidence of Doing Mathematics and Higher-Level Questioning increased substantially over the course of the project. Teacher Gathers and Synthesizes Multiple Strategies or Students Build on Other Students' Strategies increased each year and then declined slightly at time 4. Evidence of Mathematical Accuracy increased each year but declined slightly at time 4. Use of math notebooks nearly doubled and was at its highest level at time 4. The percentage of Conceptual or Deep Connections increased each year, doubling from time 1 to time 3 and increasing to its highest level by time 4. Use of Appropriate Mathematical Language was observed with increasing frequency until time 4 when it declined. English Learner students were observed communicating their ideas and reasoning verbally or in writing clearly 80% of the time by time 3. By time 4 the percentage had declined to 19%. Independent and Guided Practice were not observed with much frequency at time 1. By time 3 they were observed 73% and 80% of the time respectively. This declined sharply by time 4.

Notebooking

There was good evidence of consistency in notebooks across each class, and evidence of concept development. Use of math notebooks appeared to decline in daily usage by time 4 as a main learning tool. This finding conflicts with the results of a similar question based on observation. At time 4, the top three elements demonstrating growth were:

- Evidence of summarizing or reflection (could be quizzes or "exit tickets")
- Pages/sections consistently numbered
- Table of Contents

Formative Assessment

The Treatment Teachers demonstrated higher growth percentages from baseline to post-on the rate at which they used formative assessment to Determine Prior Learning, Provided Students with Specific Feedback, and Varied the Resources they tended to use for teaching.

Professional Development

More than 90% of participants Learned How to Utilize Existing Resources, About Additional Resources, New Techniques, and New Knowledge from each of the trainings. 100% of the teachers indicated that they Gained New Knowledge from the training. Nearly all participants found the trainings to be very/extremely valuable. Nearly all participants were inspired by their participation. The vast majority of participants indicated Increased Interest in Math and Confidence in teaching math. Participants found the Speaker Effectiveness and Practical Knowledge/techniques to be the most valuable components of professional development.

SBA – Math

The average SBA Math scale score increased each year for students with any exposure to the program relative to students with no exposure, though not to a statistically significant degree. In the English Learner specific analysis, Treatment Students with any exposure to the program demonstrated similar growth from baseline to time 4. There was no statistically significant difference.

SBA- ELA

The average SBA ELA scale score increased each year for students with any exposure to the program relative to students with no exposure to a statistically significant degree. In the English Learner specific analysis, Treatment Students with any exposure to the program demonstrated similar growth from baseline to time 4 and this was not statistically significant.

AMC

Treatment early elementary aged students who completed the AMC improved their performance by more than 10% on nearly every section (e.g. 10 Frames, Counting Objects, Number Arrangements) of the interview with pre- and post- measures. The improvements over the course of the school year were substantial. All areas improved by 10% or more with the exception of Grouping 10's.

LMT

Treatment Teachers improved their performance from pre- to post-assessment on the first section (NCOP) of the LMT. In contrast, teachers in the Comparison condition decreased from pre- to post-assessment. When comparing the Treatment Teachers and the Comparison Teachers, it should be noted that the Treatment Teachers were performing at a lower level at baseline and improved at a greater rate over the course of the year relative to the Comparison Group. These differences were statistically significant. On the second section of the LMT (GEO), the Treatment Teachers also improved substantially from pre- to post- measurement relative to the Comparison Teachers. Unfortunately, likely due to small sample size of the Comparison group, this difference did not reach statistical significance. On the third section of the LMT (PFA), the Treatment Teachers improved substantially from pre- to post- measurement relative to the Comparison Teachers. Unfortunately, likely due to small sample size of the Comparison group, this difference did not reach statistical significance. On the fourth section of the LMT (PFA repeated), the Treatment Teachers performance grew while Comparison Teacher performance declined over the same time period. This change was also not statistically significant, presumably due to small sample size.

Math Teaching Efficacy and Expectancy Beliefs Instrument (MTEEBI)

Results of the Math Teaching Efficacy and Expectancy Beliefs Instrument reveal that Treatment Teacher confidence in their ability to teach Estimation, Using Materials to Represent Problems in Multiple Ways, and Integrating Student Strategies and Ideas into their Lessons increased the most from baseline to time 4. Concepts such as "Even with appropriate instruction, most students rarely consider whether their math work makes sense," and "Even a teacher with good math teaching abilities may not help some students learn math," decreased from baseline to time 2 and then gradually crept back up over time 3 and time 4. Overall, Treatment Teachers indicated decreases in many negative beliefs about teaching and student learning and demonstrated increases in positive beliefs.

Conclusions

TEEM EL students outperformed comparison EL students in the first year for Math and ELA and also in the second year for ELA. By year 3 there appeared to be no measurable difference between groups. There are several possibilities for the lack of differentiation. They include, historical effects, dwindling sample size, and the indirect nature of the intervention (focus on teacher training). Beginning in year 2 of the study, the political climate shifted dramatically for immigrants and non-English speaking students. In other studies, we have observed decreases in performance among English Learner students. It is possible that any gains made were obscured by overall decreases in performance due to anxiety and absences from school. Additional study would be needed to ascertain if these historical effects may have impacted student performance. In addition, sample size began to decrease each year of the study making statistical significance more challenging to attain. Previous experience has also shown that finding student level effects in teacher focused projects can be difficult.

The LMT results are a reflection of teacher learning in the domain of mathematics. For all four years of the study TEEM teachers outperformed the comparison teachers. The effect sizes for all four years were above the WWC threshold for a substantial and important effect though statistical significance was only met in years 1 and 3, likely due to the sample size of the comparison group who's data was difficult to obtain each year.

TEEM teacher math efficacy and expectancy beliefs improved throughout the project and were more positive relative to the comparison teachers each year of the project, though statistical significance was likely not reached in year 4 due to small sample size of the comparison group.

The following goals were met or exceeded:

Teachers demonstrated improvement above the 30% benchmark in the quality and frequency of classroom formative assessment (3).

Early elementary students who took the AMC demonstrated substantial gains from the beginning to the end of each school year. However, since we were unable to obtain a comparison group, in our analysis there is no way to know if the gains were related to treatment exposure (1).

Treatment teachers outperformed the comparison group each year of the project on the various sections of the LMT (2).

Treatment teachers expectancy and efficacy beliefs in the domain of mathematics did decrease slightly over the course of the project, which may have been a case of the more you know the more you know what you don't know, but still outperformed the comparison group (2).

The following goals were not met:

Treatment EL students performed similarly to comparison students on the SBA-Math and ELA measures and did not meet the 5% per year improvement benchmark (1).

Observations did not demonstrate consistent implementation at the 25% improvement level (3).

References

Amaral, O. M., Garrison, L., & Klentschy, M. (2002). Helping English Learners Increase Achievement Through Inquiry-Based Science Instruction. *Bilingual Research Journal*, 26(2), 214. Retrieved from http://ehrweb.aaas.org/UNESCO/pdf/Klentschy.pdf

American Educational Research Association (AERA) (2006). Do the math: Cognitive demand makes a difference. *Research Points* 4(2). Retrieved from http://aera.net/Publications/ResearchPoints/tabid/10234/Default.aspx

Callahan, R. (2013). The English Learner Dropout Dilemma: Multiple Risks and Multiple Resources. *California Dropout Research Project Policy Brief 19*. Retrieved from http://www.cdrp.ucsb.edu/download.php?file=policybrief19.pdf

Carpenter, T., Fennema, E. and Franke, M. (1999). *Children's Mathematics: Cognitively Guided Instruction*. Heinemann

Celedón-Pattichis, S., & Ramirez, N. G. (Eds.). (2012). *Beyond good teaching: Advancing mathematics education for ELLs*. Reston, VA: National Council of Teachers of Mathematics.

City, E., Elmore, R., Fiarman, S. and Teitel, L. (2009). *Instructional Rounds in Education: A Network Approach to Improving Teaching and Learning.* Harvard Education Press

Common Core State Standards Initiative (2014). Key Shifts in Mathematics. Retreived from http://www.corestandards.org/other-resources/key-shifts-in-mathematics/

Cuoco, A., Paul Goldenberg, E., & Mark, J. (1996). Habits of mind: An organizing principle for mathematics curricula. *The Journal of Mathematical Behavior*, *15*(4), 375-402. Retrieved from http://www2.edc.org/CME/showcase/HabitsOfMind.pdf

Darling-Hammond, L et. al (2009). Professional Learning in the Learning Profession: A Status Report on Teacher Development in the United States and Abroad. National Staff Development Council. Retrieved from http://learningforward.org/docs/pdf/nsdcstudy2009.pdf

Dweck, C. (2006). *Mindset: The new psychology of success*. Random House LLC.

Enochs, L. G., Smith, P. L. & Huinker, D. (2000). Establishing factorial validity of the mathematics teaching efficacy beliefs instrument. *School Science and Mathematics*, v100 n4 p. 194-202.

Epstein, A., et. al. (2004). Preschool Assessment: A Guide to Developing a Balanced Approach. National Institute for Early Education Research Preschool Policy Matters Brief #7.

Hallinger, P., Murphy, J. (1985). Assessing the Instructional Management Behavior of Principals. *The Elementary School Journal,* Vol. 86, No. 2, p. 217-247

Hargrove, T. Y., & Nesbit, C. (2003). Science Notebooks: Tools For Increasing Achievement Across the Curriculum. *ERIC Digest*. Retreived from http://files.eric.ed.gov/fulltext/ED482720.pdf

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal*. v. 105 n. 1.

Leithwood, K., Louis, K., Anderson, S., Wahlstrom, K. (2004). *Review of Research: How Leadership Influences Student Learning*. Center for Applied Research and

Educational Improvement. The University of Minnesota. Retrieved from http://www.wallacefoundation.org/knowledge-center/school-leadership/key-research/Documents/How-Leadership-Influences-Student-Learning.pdf

Lewis, C. (2004) Does Lesson Study Have a Future in the United States? JSSE Vol. 3, No. 1. Retrieved from www.jsse.org/index.php/jsse/article/download/967/870

Lewis, C., Perry, R, Hurd, J. (2009). Improving Mathematics Instruction through Lesson Study: A Theoretical Model and a North American Case. *Journal of Mathematics Teacher Education*, vol. 12, no. 4, p 285-304. Retrieved from http://elac.ex.ac.uk/lessonstudymld/userfiles/Lewis%20et%20al%202009.pdf

Liptak, L. (2005). For Principals: Critical Elements. In *Building Our Understanding of Lesson Study*, Wang-Iverson, P. and Yoshida, M., Eds. p.39-44.

Moschkovich, J. (2013). Principles and Guidelines for Equitable Mathematics Teaching Practices and Materials for English Language Learners. *Journal of Urban Mathematics Education*, Vol. 6, No. 1, pp. 45-47. Retrieved from http://ed-osprey.gsu.edu/ojs/index.php/JUME/article/download/204/135

National Mathematics Advisory Panel. (2008). *Foundations for Success: The Final Report of the National Mathematics Advisory Panel.* U.S. Department of Education.

National Research Council. (2001). *Adding it up: Helping children learn mathematics.* In J. Kilpartrick, J. Swafford, & B. Findell (Eds), Washington, DC: National Academy Press.

Parrish, S. (2010) *Number Talks, Grades K-5: Helping Children Build Mental Math and Computation Strategies*. Math Solutions.

Richardson, K (2011). Assessing Math Concepts Assessments. Math Perspectives

Riggs, I. M., Fischman, D. D, Riggs, M. L., Jetter, M. E., & Jesunathadas, J. (2018). Measuring teachers' beliefs in relation to teaching mathematics with mathematical practices in mind. *School Science and Mathematics*, vol. 118, no. 8 pp. 385-395.

Rumberger, R., and Lim, S. (2008). Why Students Drop Out of School: A Review of 25 Years of Research. *California Dropout Research Project Research Report 15.* Retrieved from http://www.cdrp.ucsb.edu/download.php?file=researchreport15.pdf

Shimojyo, H & Shimojyo, Y. Science Notebooks: A Powerful Learning Program. Presentation to *MaTHink* annual mini-conference, Riverside County Office of Education, March 2013. Retrieved from http://algebraforum.wikispaces.com/file/view/Shimojyo+Notebooking+Powerpoint.pdf

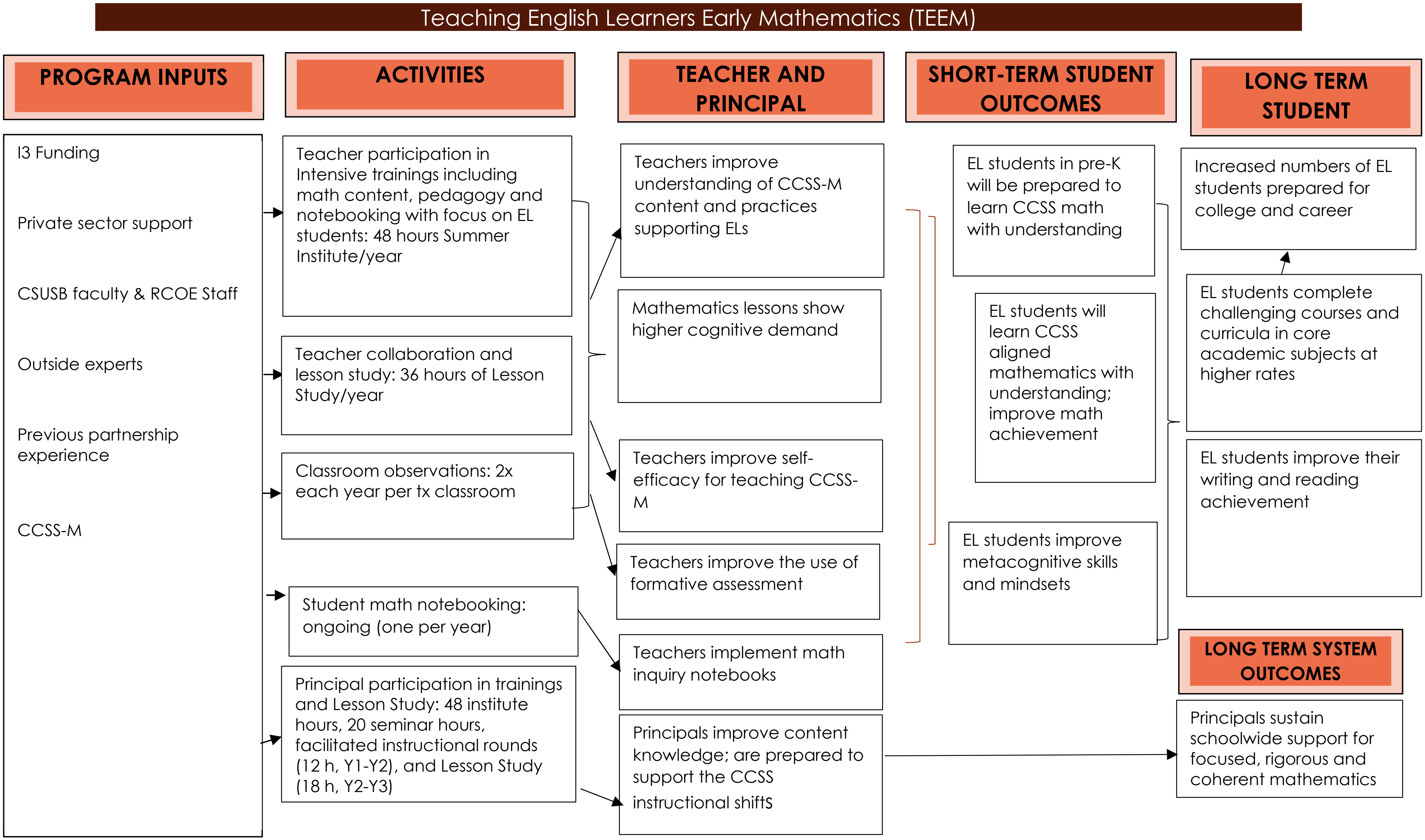
Silver, D., et al. (2008). What Factors Predict High School Graduation in the Los Angeles Unified School District? *California Dropout Research Project Research Report 14*. Retrieved from http://www.cdrp.ucsb.edu/download.php?file=researchreport14.pdf

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. Simon and Schuster.

Terry, L. and Rosin, M. (2011). California's Math Pipeline: Success Begins Early. *EdSource Issue Brief.* Retreived from http://edsource.org/wp-content/uploads/STEM-EdSourceBrief-MathPipeline072211.pdf

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, v. 17,* pp. 783-805. Retrieved from http://anitawoolfolkhoy.com/pdfs/aeraefficacy-measure.pdf

Appendix 1 Logic Model



Teaching English Learners Early Mathematics (TEEM)

PROGRAM INPUTS

I3 Funding

Private sector support

CSUSB faculty & RCOE Staff

Outside experts

Previous partnership experience

CCSS-M

ACTIVITIES

Teacher participation in Intensive trainings including math content, pedagogy and notebooking with focus on EL students: 48 hours Summer Institute/year

Teacher collaboration and lesson study: 36 hours of Lesson Study/year

Classroom observations: 2x each year per tx classroom

Student math notebooking: ongoing (one per year)

Principal participation in trainings and Lesson Study: 48 institute hours, 20 seminar hours, facilitated instructional rounds (12 h, Y1-Y2), and Lesson Study (18 h, Y2-Y3)

TEACHER AND PRINCIPAL

Teachers improve understanding of CCSS-M content and practices supporting ELs

Mathematics lessons show higher cognitive demand

Teachers improve self-efficacy for teaching CCSS-M

Teachers improve the use of formative assessment

Teachers implement math inquiry notebooks

Principals improve content knowledge; are prepared to support the CCSS instructional shifts

SHORT-TERM STUDENT OUTCOMES

EL students in pre-K will be prepared to learn CCSS math with understanding

EL students will learn CCSS aligned mathematics with understanding; improve math achievement

EL students improve metacognitive skills and mindsets

LONG TERM STUDENT

Increased numbers of EL students prepared for college and career

EL students complete challenging courses and curricula in core academic subjects at higher rates

EL students improve their writing and reading achievement

LONG TERM SYSTEM OUTCOMES

Principals sustain schoolwide support for focused, rigorous and coherent mathematics

Core theories and assumptions: Research on Mathematical habits of mind (Cuoco et al, 1996), cognitive demand as a feature of effective instruction (AERA, 2006), growth mindset (Dweck, 2006), teacher and student self-efficacy (Tschannen-Moran, Hoy, 2001).

Appendix 2 Fidelity of Implementation Matrix

TEEM: Fidelity of Implementation Table

Component	Operational Definition of Indicator	Unit of Implementation	Data Source for Measuring Indicator	Individual Level Metric	Scoring at Level 1 (Teacher, student, principal)	Level 1 Threshold for Adequate Implementation	Program Level Threshold for Adequate Implementation	Expected Sample	Years of Fidelity Measurement
Teacher participation in trainings	6 full days of training per year, 48 total hours	teacher	Attendance records	Days of training	0 = Teacher attended less than 5 full days (or < 32 hours) of training; 1 = Teacher attended at least 5 but less than 6 full days (or 40 hours) of training; 2 = Teacher attended 6 full days (or 48 hours) of training	high implementing teacher = score of 2	Across the entire sample, at least 80% of TEEM teachers will have score of 2	74 teachers, i.e., all TEEM teachers	2015-16; 2016-17; 2017-18; 2018-19
Teacher collaboration	36 hours/year	teacher	Attendance records	Number of hours	0 = Teacher attended less than 28 hours of collaboration time; 1 = Teacher attended at least 28 hours but less than 36 hours of collaboration time; 2 = Teacher attended 36 hours of collaboration time	high implementing teacher = score of 2	Across the entire sample at least 80% of TEEM teachers will have score of 2	74 teachers, i.e., all TEEM teachers	2015-16; 2016-17; 2017-18; 2018-19
Classroom observations	2x per teacher, baseline year; 1x each year thereafter	teacher	Classroom Observation Tool maintained by Staff Developers	Number of observations	0 = 59% of participating teachers observed; 1 = 60% to 79% of participating teachers observed; 2 = 80% to 100% of participating teachers observed	high implementing teacher = score of 2	Across the entire sample at least 80% of TEEM teachers will have score of 2	74 classrooms, i.e., all TEEM classrooms	2015-16; 2016-17; 2017-18; 2018-19
Principal participation in trainings	50 total hours	principal	Attendance records kept by developer	Number of hours	0 = Principal attended less than 30 hours of collaboration time; 1 = Principal attended at least 30 hours of collaboration time but less than 40 hrs; 2 = Principal attended at least 40 hours of collaboration time	high implementing principal = score of 2	Across the entire sample at least 5 TEEM principals will obtain a score of 2	7 principals, i.e., all TEEM principals	2015-16; 2016-17; 2017-18; 2018-19
Student Notebooking	Percent essential elements included	One notebook per student during the year	Students' notebooks	Student-level score on their notebooks based on the rubric (i.e., that lists essential elements)	0 = less than 60% rubric elements included 1 = at least 60% but less than 80% rubric elements; 2 = at least 80% of rubric elements included	high implementing student = score of 2	Of a random sample of notebooks, at least 80% of the students will obtain score of 2 on their notebooks	222 randomly selected student's notebooks, 3 from each TEEM classroom	2015-16; 2016-17; 2017-18; 2018-19

Appendix 3

Impact Contrast Table

	Treatment Group N of Students or Teachers	Comparison Group N of Students or Teachers	Treatment Group SD	Comparison Group SD	Comparison Group Mean	Impact Estimate	Standardized Effect Size Hedge's g	Impact Standard Error	Degrees of Freedom	Exact p-value
Smarter Balanced Assessment - Math										
Year 1	541	625	93.76	68.65	2434.82	25.08	0.15	6.45	590	0.00
Year 2	207	723	81.15	74.41	2428.75	1.583	0.23	7.09	256	0.82
Year 3	187	535	72.97	70.57	2414.33	1.604	0.17	5.48	362	0.77
Year 4	308	566	93.31	96.95	2480.81	-12.14	0.16	11.57	183	0.29
Smarter Balanced Assessment – English Language Arts										
Year 1	539	710	99.74	87.37	2447.84	23.66	0.09	6.44	423	0.00
Year 2	207	808	81.80	86.57	2426.08	14.28	0.38	6.57	140	0.03
Year 3	187	619	82.81	81.90	2405.98	-6.154	0.22	5.19	359	0.24
Year 4	305	554	80.09	87.45	2395.60	-10.85	0.20	10.98	361	0.32
Learning Mathematics for Teaching										
Year 1	62	40	0.802552	0.717419	-0.19736	0.27	0.33	40.292	101	0.04
Year 2	34	19	0.7856406	0.5095198	-1.146085	0.24	0.36	17.329	52	0.16
Year 3	42	33	0.707877	0.915931	-0.52838	0.38	0.35	0.18	72	0.04
Year 4	33	14	0.804872037	0.8251801	-0.372142	0.33	0.28	0.247	44	0.19
Math Teaching Efficacy and Expectancy Beliefs Instrument										
Year 1	56	31	9.44689	9.6076	78.0	4.64	0.17	1.386	74	0.001
Year 2	21	11	8.29357	7.61555	76.9	6.09	0.46	2.833	23	0.04
Year 3	31	23	8.12576	8.44136	80.8	6.19	0.36	2.045	39	0.004
Year 4	32	9	4.60498	3.65874	69.8	0.43	0.29	2.886	20	0.884